NWS Operations Proving Ground

Final Report

Real-Time Severe Convective Weather Experiment

An OPG Capacity Testing Experiment Focused on the Mesoscale Environment

NWS Operations Proving Ground
7220 NW 101st Terrace
Kansas City, MO 64153

1. Executive Summary

Thanks to the procurement of AWIPS in the Cloud in late 2020, in May, 2021, the OPG attempted a rather ambitious real time demonstration of a mesoscale analysis approach for WFO operations. We designed the demonstration to answer questions related to 7 objectives based upon the results of the 2019 Mesoscale Bootcamp.

61 participants representing 44% of WFOs within the CONUS joined the OPG in two day sessions over the course of two weeks. In addition, 16 subject matter experts served in a “mutual-aid” type role and assisted each of the participants in diagnosing the mesoscale environment.

The participants used AWIPS in the Cloud, and other web-based data sources, to interrogate the environment. Then, using ArcGIS Online (AGOL), the participants drew polygons on a map representing an “area of concern” (AC) in the one to six hour valid period. They were instructed to draw an AC when any of the following conditions were anticipated: new convection, severe convection (including tornadoes), and improving conditions (decreases in convective activity).

The OPG evaluated the participants' AC polygons to determine skill, evaluate the impact of convective allowing models (CAMs) on the forecast process, assess the value of geospatial collaboration, and evaluate the impact of probabilistic messaging.

In hindsight, we now realize our goals were likely too ambitious. Our experimental design, while well thought out, produced mixed results for some objectives. For example, we simply can not say for certain whether forecasters demonstrated skill in assessing the mesoscale environment. We learned several valuable lessons that made this experiment a worthwhile experience for the OPG our participants.

We showed, without question, that a large number of forecasters from around the country with diverse backgrounds can interact effectively using cloud based technologies to perform highly complex meteorological analysis in real time (all from the comfort of their homes). We showed that geospatial collaboration tools are a key to unlocking the full potential of any collaborative forecast process. We showed that the mutual aid approach to mesoscale analysis is not only effective, but it is viable, achievable, and can be completed by forecasters from around the country.

We learned several valuable lessons about experimental design within the context of a large scale virtual demonstration. We developed a geospatial method for evaluating forecaster skill in various tasks and we believe we have a structure for assessing the value of certain datasets in the forecast process.

We uncovered information about the inefficient nature of forecast operations that, while not related to our original objectives, are crucial to understanding the challenges of adopting mesoscale analysis in operations today.

This report will provide details on the background of our experiment, our results, and of course findings and recommendations. We have also designed a special **ArcGIS Online "Story Map"**[1] to provide a more interactive reader experience.

2. Background (Spirit and Intent)

We know that high quality, targeted, and consistent Impact-Based Decision Support Services (IDSS) during short term hazardous weather events requires skillful mesoscale analysis (hereafter called 'mesoanalysis')(e.g., Togstad et al. 2011; Gravelle et al. 2016a, b).

Over the last few years, some local offices have implemented newer operational structures with a Mesoanalyst role clearly defined and staffed during severe events.

> *"My office already utilizes a very similar workflow with open discussion between the Mesoanalyst, IDSS/communications, radar operators, etc... This works well for our severe operations and everyone is on the same page in terms of expected convective evolution and ongoing trends."* — Mesoscale Experiment Participant

Unfortunately, for a host of reasons, the Mesoanalyst role is still not ubiquitous throughout the agency.

> *"We are still trying to achieve full staff buy-in for staffing a dedicated Mesoanalyst for every event."* — Mesoscale Experiment Participant

Through many conversations with regional headquarters, local offices, national program leads, and national centers, the OPG believes the operational community overwhelmingly supports integrating dedicated mesoanalysis in operations. However, there are still some lingering questions regarding the gaps that exist in NWS products and services and how best to fill those gaps in the short term forecast timeframe.

The primary purpose of this OPG experiment was to follow up on findings from the 2019 "Mesoscale Bootcamp" (MBC). In 2019, the OPG co-hosted several interactive workshops with SPC, OCLO, and NSSL focusing on improving skill and application of mesoanalysis techniques during convective events. The MBC was a resounding success, but the final report from the bootcamps indicated a few challenges the NWS needs to overcome to fully realize the value mesoanalysis can provide. The following is an abbreviated list of those findings:

1. Skillful mesoanalysis, coupled with proactive communication, has the potential to enhance multiple aspects of NWS service delivery.
2. Training will be required in order to build the knowledge and skills needed to perform skillful mesoanalysis.
3. Foreseeing a spectrum of possible scenarios in advance can prevent forecasters from focusing on a single particular outcome.

[1] The StoryMap requires an LDAP username and password to access - click HERE to view log in instructions.

4. Determining and communicating the risk of severe weather lends itself well to the introduction of probabilistic information.
5. Cultivating an office culture characterized by a general openness to others' perspectives, a continual exchange of ideas, and a deliberate engagement with stakeholders to learn what is important to them, is critical to translating skillful mesoanalysis into effective IDSS.
6. Mesoanalysis is often most valuable in situations where the threat of organized severe storm development is highly conditional.
7. Some offices simply do not have the staffing and/or on-site expertise to take advantage of the value a skilled Mesoanalyst might provide for local high-impact events.
8. Adopting the practice of assigning a dedicated Mesoanalyst in WFO operations offers an attractive avenue of meaningful contribution for those forecasters whose skills and interests lie in providing the deep science necessary to support colleagues who are crafting and delivering IDSS messaging.
9. By expanding the knowledge and skills associated with effective mesoanalysis, and adopting a more SPC-like anticipatory mindset concerning the evolution of convective events, WFO-SPC collaborations can be enriched, with mutual benefit.

Since 2019, the NWS has addressed a few of the challenges. For instance, there is now a comprehensive web based training program to improve mesoanalysis capabilities in addition to AWOC training materials. Local offices are integrating mesoanalysis into operations more frequently as an integral component of severe events. Regional offices are conducting experiments focusing on collaboration with SMEs in real time. Finally, SPC is working diligently to ensure local offices are aware and active participants in collaboration opportunities.

As a natural extension of the 2019 Mesoscale Bootcamp, the OPG, in collaboration with the AFS Severe Weather Program, designed a virtual mesoanalysis applications experience for 61 NWS forecasters. This real-time experience focused on integrating intra- and inter-WFO geospatial collaboration with SMEs, a probabilistic perspective on risk assessment, and a scenario-based IDSS communications concept. Thus, the primary spirit and intent of our experiment was to simply apply lessons learned (recommendations) from the Mesoscale Bootcamp in real-time and evaluate the results.

3. Experimental Design

61 participants (60 meteorologists and one ITO) and 16 SMEs joined the OPG over a two week period in late May, 2021 for a real time severe convection related experiment (Fig. 1).

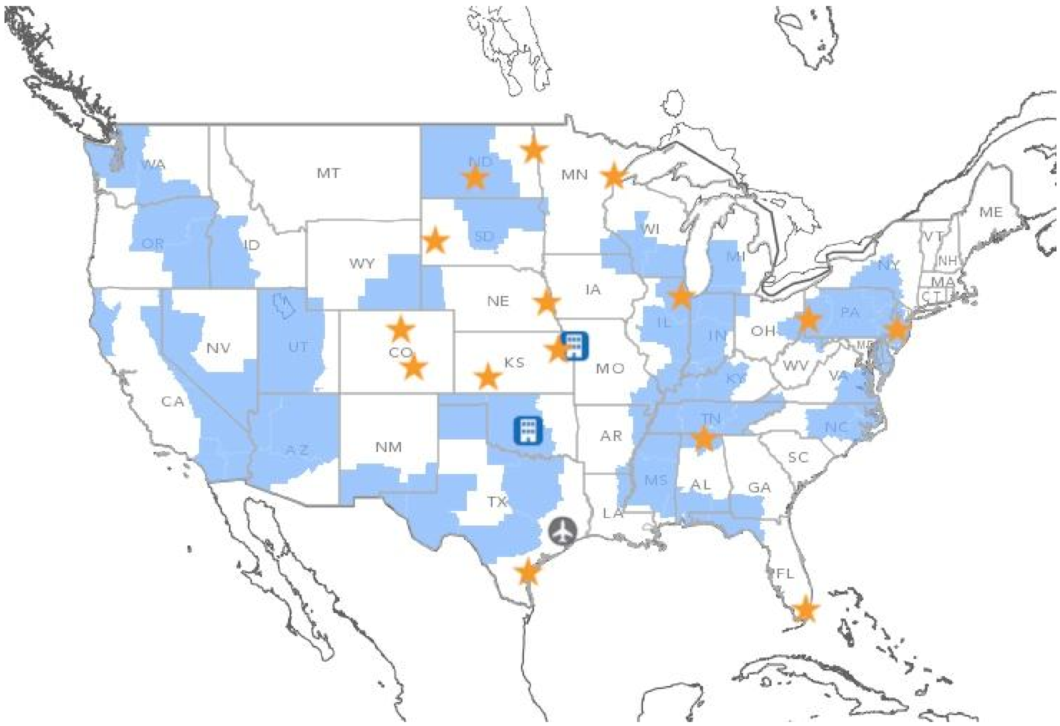

Figure 1: Map of participant office locations (blue shading), SME locations (stars), NCEP center participants (blue icons with buildings), and CWSU participant (airplane icon).

We conducted the experiment in two-day chunks with roughly sixteen participants and eight SMEs in each chunk. Two participants joined each other as a "shift team" each day representing a WFO under threat of deep convection or severe weather during the afternoon and evening hours. Each shift team was joined by an SME who led the discussion about the mesoscale environment on the first day, then served as a trusted advisor on the second day.

The two-person shift teams, and their assigned SME, worked collaboratively in a Google Meet session and leveraged a Google Chat room to interact with "neighboring WFOs" (the other experiment participants). Through their analysis, and collaboration, the shift teams produced information in two formats: a polygon drawn using AGOL, and Google Slides.

Upon drawing a polygon in AGOL, participants had to input information into a pop up window including:

- Starting valid time of the polygon (at least 1 hr into the future)
- Ending valid time of the polygon (no more than 6 hr into the future)
- Most likely conditions expected
- Worst case condition expected
- Best case conditions expected
- Forecast Discussion (brief science-based reasoning for the polygon)

The polygon would then become visible to all of the experiment participants, SMEs, and observers in AGOL in real time. Further, if a participant wanted to modify the polygon, they could do so even if they did not originally create the polygon. However, the OPG made it clear that the participants should only modify polygons as part of the collaborative process, but once all parties agreed on the location and content of the polygon, no further edits or updates should be made.

Our participants also created IDSS content using Google Slides to simulate PowerPoint briefings, and Google Docs to simulate partner emails, convective Area Forecast Discussions (AFDs), or NWS Chat messages. The Google suite of products allowed participants to collaborate on their IDSS content in real time.

3.1 IDSS Requests

Throughout the experiment, the OPG would provide “injects” to our participants through AGOL. Participants would see a red dot appear on their editable maps. The red dots contained information supplied by the OPG referencing various IDSS requests. Some of the requests vere vague and simply asked for “*the condition expected in a few hours*”, while other requests were specific, “*what are the chances we see wind gusts above 45 mph?*”. Participants could then enter a reply through AGOL which generated a little NWS logo on their shared map (Figures 2 and 3).

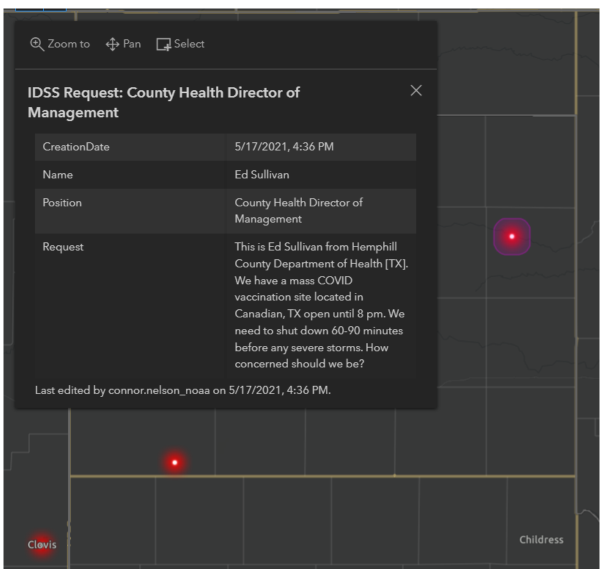

Figure 2: Example of IDSS inject requests (red dots) with accompanying text. The request text reads: *“This is Ed Sullivan from Hemphill County Department of Health. We have a mass COVID vaccination site located in Canadian open until 8 pm. We need to shut down 60-90 minutes before any severe storms. How concerned should we be?”*

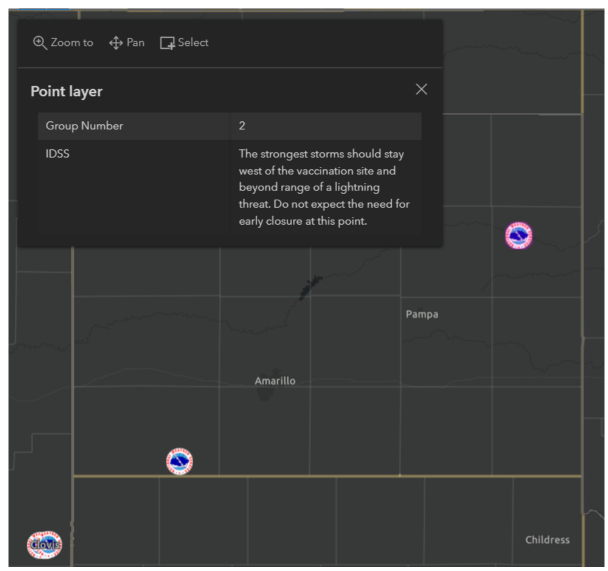

Figure 3: Participant responses (NWS logo) to the IDSS inject requests in Figure 1 and accompanying text. The IDSS response reads: *“The strongest storms should stay west of the vaccination site and beyond range of a lightning threat. Do not expect the need for early closure at this time.”*

The OPG did not verify the accuracy of the IDSS responses. Instead, we reviewed the content for common themes (discussed further in section 4 below) and discussed the concept during debriefings.

3.2 Data Denial Component

One of the challenges facing forecasters in operations today is the overwhelming amount of data that can be integrated into the forecast decision-making process. Forecasters have access to a wealth of observational data, real time analysis data, and of course numerical model data in both deterministic and ensemble frameworks. The OPG does not contend that we know which data are most valuable or least valuable to use during mesoscale operations, so we instead attempted to assess the impact on the forecast process when certain data were restricted.

As such, half of the experiment participants were asked to avoid using convective allowing models (CAMs) of any kind (deterministic, ensemble, operational, experimental, etc). They were still allowed to review analysis data (such as the SPC mesoanalysis page) and global models though.

3.3 Roles and Responsibilities

Each day of the experiment, the shift teams picked their roles. One participant would serve as the primary collaborator with neighboring offices and would draw the polygons in AGOL. The second participant served as an "IDSS Communicator" who developed the IDSS materials described previously. Our participants often switched roles between the first day and second in order to experience both aspects of the experiment.

On the first day of each experiment, the SME facilitated the mesoanalysis. They provided their expert opinion, and effectively led the conversations. On the second day of the experiment, the SME served in a more passive role allowing the participants to lead the analysis and collaboration. The SME was able to interject or provide assistance upon request. In this way, our SME served in both a de facto "mutual aid" capacity and a teaching capacity.

3.4 Pre-work and Training

Prior to the experiment, the OPG asked all participants to complete a few of the Mesoscale Environmental Analysis (MEA) modules including:

Satellite Imagery Analysis	Targeting Tornadic Threats	Targeting Hail & Damaging Winds
Mesoanalyst Operational Workflow	Risk Management in Severe Wx Ops	IDSS Messaging

3.5 Participant Diversity and Inclusion

The OPG believes strongly in designing and selecting participants who represent the diversity of operations. In this case, we wanted participants with various degrees of mesoanalysis skill, operational experience, and geospatial diversity. By “going virtual”, we are able to accomplish these goals more effectively than in-residence experiments.

Our 61 participants represented 46 WFOs and one NCEP center (AWC). Our 16 SMEs represented 11 WFOs, one NCEP center (SPC), one CWSU and the NWSTC. So combined, our experiment represented 51 WFOs or roughly 44% of all CONUS WFOs. We also had a diverse mix of positions and years of experience (Figs. 4 and 5).

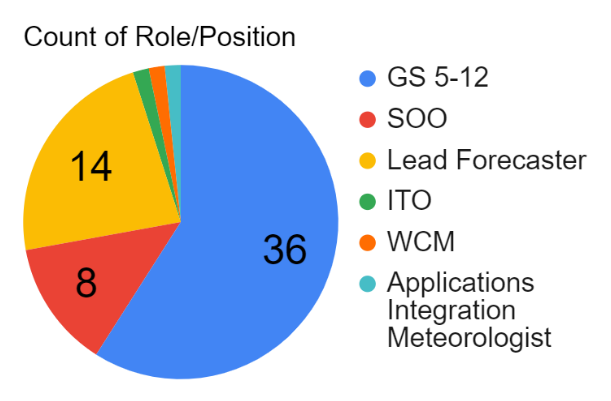

Figure 4: Pie chart displaying current job position of the participants.

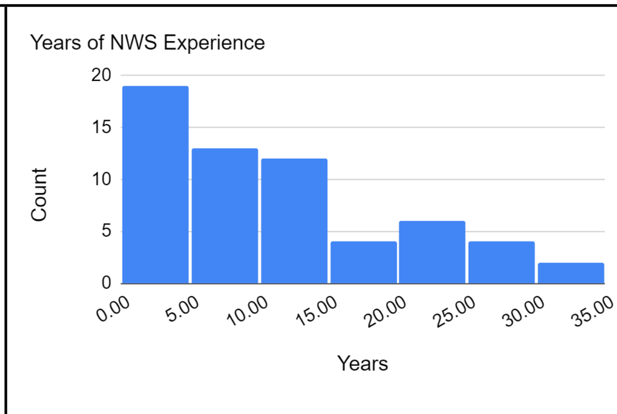

Figure 5: Histogram of the experience distribution of participants

Further, fourteen women joined as either participants (11) or SMEs (3), which is roughly 19% of all individuals who took part in the experiment.[2]

4. Objectives and Findings

Our broad goal in this experiment was to understand how collaborative structures could be applied in operations today to maximize the mesoanalysis process. The information below describes each of the original objectives for the experiment followed by a finding for each objective. In section 6, we summarize the findings and present a few recommendations based on a holistic view of the findings.

Objectives 1 and 2 focus on participant skill and the tools used to analyze the mesoscale environment. Objectives 3, 4, and 5 all relate to the collaborative process. They address specific aspects of collaboration from the use of geospatial collaboration tools (Objective 3), to the chains of communication (Objective 4), and the value of assigning the role of “IDSS Communicator” (Objective 5). Leveraging AGOL allowed participants to focus their collaborative efforts on tangible deliverables (a

[2] On a rather odd note, eight of our participants were named Mike or Michael, who represented 9% of all the Mike’s or Michael’s in NWS operations at the time of the experiment.

polygon on a map). The SME who joined each WFO shift team served in a facilitator capacity at times to ensure conversations remained focused on the mesoscale environment. Finally, assigning clear roles and responsibilities to each participant helped the participants remain focused on their specific duties.

Finally, Objectives 6 and 7 relate to the participants capacity to assess and present probabilistic information within the context of short term, dynamic events.

Each of these objectives was designed to address the findings, noted in section 2 above, from our 2019 Mesoscale Bootcamp. We hope the findings and recommendations from these objectives inspire conversations and eventually translate to improved services during convective events.

4.1 Objective 1: Evaluating Skill

In presenting the participant polygon based verification, the OPG was neither arguing for or against local offices producing such polygons in real operations. We asked participants to create polygons with mesoscale information so we could capture their reasoning at a specific moment in time for a defined area.

The OPG understands that many local offices are expanding their use of mesoanalysis to inform short term IDSS and warning decisions. The verification can therefore serve as a baseline for future discussions such as: “how can the Mesoanalyst best support operations'', or, “does remote mesoanalysis support positively impact local operations?”

Objective #1: Evaluate forecaster skill in predicting ACs in the one to six hour time frame.

Spirit and intent: The OPG wanted to understand if forecasters could skillfully identify AC, and in theory translate that skillful information to partners in the form of enhanced IDSS.

Hypothesis: By leveraging quality mesoanalysis, well-structured collaboration techniques, and support (mutual aid) from SMEs, our participants will demonstrate valuable skill in predicting the potential for convective warnings or severe convection in the one to six hour timeframe.

Evaluation Method: To assess forecaster skill, the OPG first had to determine what the participants were predicting when they drew a polygon. We categorized polygons based on the content within the most likely field. We then determined whether the participant predicted “new convection”, “severe convection”, or “improving conditions”[3]. Table 1 below shows three examples of “most likely” text fields and the associated OPG categorization:

[3] We originally wanted to verify polygons predicting “new convection” or “improving conditions” but found the effort to be more difficult than anticipated. After verifying the “severe convection” polygons, we felt we had enough data to support our findings.

Table 1: Example text from the “Most Likely” fields of polygons and how the OPG categorized that text.

Text in “Most Likely” Field	**OPG Categorization**
“Multiple supercells with large hail up to 2.5 inches, damaging winds and a couple of tornadoes.”	We considered this a “severe” polygon.
“Once the cap erodes, expect storms to develop across the SE portion of our area along the leading edge of a moisture surge.”	We consider this a “new convection” polygon.
“Convection along a surface trough continues to progress eastward, and will exit the CWA by 2300 UTC, ending the threat for severe weather.”	We consider this an “improving conditions” polygon.

After categorizing all 280 polygons produced by our participants during the experiment, we determined 163 polygons predicted severe conditions.

We used official convective warnings issued during our experiment along with LSR reports as the verification sources. We chose warnings as a verification source based on the end user decision making process. We know partners take some action when a convective warning is issued. However, some partners need time to prepare for their actions: more time than is currently allowed by convective warnings alone. So if forecasters can predict the potential for convective warnings on a smaller scale than Convective Watches, then we can effectively give partners more time to prepare for various life saving actions.

We performed the following steps[4] to verify each severe polygon:

1. We identified all the convective warnings that aligned with our participant drawn AC polygons in both time and space.
2. We likewise identified all of the LSRs issued that aligned with the AC polygons in both time and space.
    a. We assigned a 15-mi radius buffer (based on published research analyzing the heterogeneity of near-storm environments by Dr. T. Connor Nelson; Nelson et al. 2021) to each official storm report.
3. We then compared the convective warning areas and the buffered LSR areas to the AC polygons to compute:
    a. The percent area overlap
    b. The lead time from when the AC was created to when a convective warning was issued
    c. The “False Alarm Area” which is where an AC was drawn but no warning or LSRs occurred.

[4] We also calculate the percent of polygons that experienced a severe LSR. However, due to the size of the polygons, it was difficult to infer skill based on the metrics. Very large polygons with one LSR, for example, would be counted in the verification even though most of the polygon area likely experienced no severe convection.

Next, we needed to establish a baseline for comparison in order to contextualize the resulting values. We decided to use convective watches for the baseline because our goal was to evaluate if forecasters can provide valuable IDSS after convective watches are issued. SPC had issued 15 convective watches covering our experiment domains and valid times (It turns out that no watches were issued on May 19th and May 20th).

We calculated the same Percent Area Coverage and False Alarm Area statistics for the convective watches that we computed for the polygons. We found that, on average, a combination of official warnings or buffered LSR reports covered about 50% of the convective watches issued.

To ensure the values were comparable, we needed to consider only the AC polygons that were issued within a convective watch. That left us with 127 participant drawn AC polygons that aligned in time and space with convective watches. All of the verification we present on AC polygon verification and thus based on those 127 polygons that predicted severe conditions and fell within a convective watch.

Results: We found that, on average, a combination of buffered LSRs and convective warnings covered about 35% of the AC polygons produced by our participants (considering only the 127 AC polygons that overlapped with convective watches). The following notched box and whisker plot shows the AC polygon verification (percent of polygon area covered by warnings or LSRs) by experiment day (note that there is no data for May 19th and 20th because no watches were issued those days):

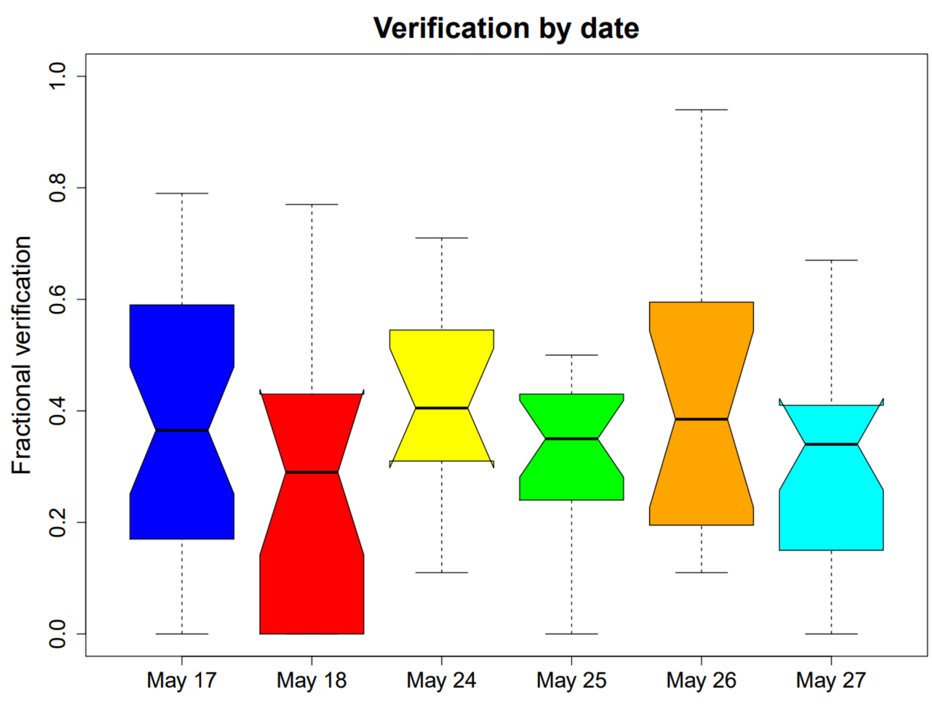

Figure 6: Notched box plots of fractional verification (scale of 0 - 1) on each day of the experiment (colored by date). The thick solid black line is the median of the distribution and the notches are representative of the 95% confidence level about the median.

When examining the distribution of verification performance each day of the experiment, we see that the median, interquartile range limits, and whisker values generally drop from “Day 1” (first day of each session, where SMEs lead the analysis) to “Day 2” (the second day of each session, where SMEs were ancillary) by approximately 6% (Fig. 6). A positive causal relationship between the presence of SMEs and verification or impact on mesoanalysis is PROBABLE, but given the issues in subjectiveness of polygons as they relate to 'severe' and WEPs and the total number of data we cannot definitively make a quantitative statement on the value. We can, however, show that forecasters felt positive about a mutual aid or SME approach and had an effect on their "confidence". Thus, further testing needs to be done to quantify the relationship and its value under more controlled experimental conditions.

4.1.1 Event Magnitude and Impact on Verification

We thought that event magnitude may have played a role in the resulting verification. Table 2 below shows the number of warnings and LSRs issued within our domains during the experiment.

Table 2: The number of warnings and LSRs that occurred on each data of the experiment along with the median percent of polygon area that “verified”

Date	**Number of Warnings**	**Number of LSRs**	**Median % Area Verified**
May 17th	108	91	37%
May 18th	27	32	29%
May 24th	53	44	41%
May 25th	22	60	35%
May 26th	65	119	38%
May 27th	96	125	34%

We also ran a Matrixed Student’s T-Test to compare the verification from all the experiment dates and found the only statistically different data sets were May 18th and May 24th. Given that only two dates were statistically different even though we had a diverse number of warnings and LSRs each day, we concluded that the event magnitude had little to no effect on the resulting verification.

4.1.2 Location of AC and Impact on Verification

We also investigated the impact of location on verification. We thought that polygons covering high population density locations (urban areas) would verify better than polygons with low population density (rural areas). It turns out, there was no significant correlation between population density and polygon verification (Fig 7).

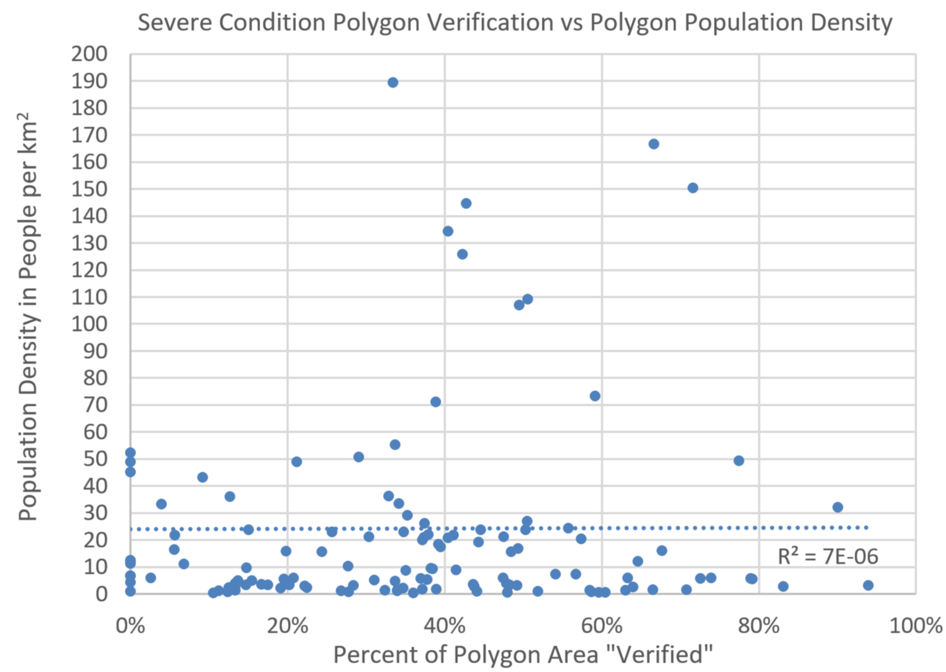

Figure 7: Scatter plot of “percent of polygon area verified” (x-axis) and the population density of each polygon (y-axis).

In this case, our correlation coefficient (R^2) was nearly 0 indicating population density of the participant drawn polygons were not correlated with the resulting verification.

We also considered evaluating whether or not participant experience or their home region impacted verification. However, because our participants operated as a collaborative pair, we can not draw conclusions based on the individual who drew the polygon.

After running all the verification numbers we were unable to determine why some participants performed better than others. We even reviewed the content in the discussion field of polygons with greater than 50% verification and found no obvious differences among polygons with less than 50% verification.

The OPG, quite frankly, is unsure whether or not we could consider the AC polygons produced by our participants as “skillful”. Our results were, admittedly, mixed. 14 polygons (11% of the total) essentially received no verifying events meaning they produced a 100% false alarm area. On the other hand, 30 of 127 polygons produced verification at the same or greater rate than SPC watches (over 50% area covered). We could not identify any physical reason explaining why some polygons verified better than others with the one possible exception being the inclusion of the SME as a primary contributor to the polygons. We therefore offer the following factors that may have contributed to the mixed results:

1. Experimental design. We asked participants to create polygons at least once per hour, which may not be realistic based on the magnitude of the event.

2. Newness of the activity. Our participants had a wide range of operational experience. For many participants, this experiment provided their first realistic attempt at diagnosing the mesoscale environment in order to provide IDSS in a probabilistic framework.
3. NWS Experience. The OPG does have access to the participants' position title and years of NWS experience, but we could not relate that data to polygon verification because each polygon was based on a collaborative process. Therefore we do not know if experience played a role in verification.
4. Words of Estimative Probability (see section 4.1.2 for context). Participants often relied on hedge terms that made characterizing their prediction difficult for the OPG.

4.1.3 Participant Feedback on the Experiment

In addition to assessing skill, we wanted to understand how participants felt about their experience and whether mesoscale IDSS would be valuable to their local partners:

> *"I know our partners would greatly appreciate this type of information and we need to break the habit of being focused solely on warnings and nowcasting radar."*

> *"If staffing is available, then the meso/DSS collaboration seems like an optimal approach for handling convective scenarios in office."*

> *"The Mesoanalyst is a key component of convective operations and needs to be utilized more."*

> *"This was a huge eye-opening experience to the possibilities of mesoanalysis and near-term forecasting contributing to operations. The amount of information that can supplement both warning decisions and IDSS is invaluable."*

The most common argument against applied mesoanalysis in operations related to staffing and culture rather than perceived value of the information:

> *"The benefits of such a workflow model are that clear roles and responsibilities are defined and that no one person's situation awareness becomes clouded; however, the challenges are that we have a diverse mix of forecasters, ranging from very experienced to new, and that staffing and routine duties (grids, aviation, upper air) forces us to multitask more than I would like."*

> *"I think we really need to look at making better efforts in providing near term DSS based on the mesoscale environment. Normally we would simply offer generalized DSS for severe weather (risk, general timing, etc.) but I think we can take it to the next level and offer more information. We would need significant training and it would require dedicated Mesoanalysts, but it could be very valuable."*

> *"The key here is minimizing operational impact. There is a significant sensitivity to adding more into operations beyond what is already asked. So in order for this to catch on, in addition to the collaborative improvement, it would need to be a net-zero in added workload. This is where the*

CR program is very helpful, but I have several questions on how that is tactically run in support of individual offices."

Interestingly, participants offered suggestions for reducing the staffing concerns noted above. Several participants recognized the experiment itself demonstrated capacity to conduct mesoanalysis from their home while others promoted the idea of remote support by SMEs:

"This could be challenging in a short-staffed office, but it could be remedied by having outside telework support in that role."

"Teleworkers may be able to help with this more efficiently and quickly since they can start working almost instantaneously from home instead of having to get ready and come into the office."

4.1.4 Words of Estimative Probability

There is a large body of research from social scientists across multiple disciplines[5] that indicates using words as a replacement for numerical probabilities causes confusion and is prohibitive to effective decision making. One subset of this research, compiled and conducted by Dr. Joseph Ripberger and his team at the University of Oklahoma's Center for Risk and Crisis Management on the subject of risk communication, suggests that Words of Estimative Probability (WEPs) are particularly damaging to understanding risk for hazardous weather conditions.

WEPs are terms used to describe uncertainty without including numeric probabilities. Examples of WEPs include: possible, may, could, should, likely, chance, or probable. The OPG agrees with these studies and strongly believes that the use of WEPs in IDSS has negative consequences for weather literacy and interpretation of IDSS. This confusion in interpretation also manifested in our experiment. For example, the "most likely" scenario field of one of our participant's polygons stated:

"*Storms forming already in the area **will likely** tap into the good instability and form tall updrafts. Very large hail **may be possible** and with multiple rounds of storms **possible** in the area, flooding **may** become an issue.*" (Emphasis added)

The OPG thus struggled to assess the true nature of such predictions. In theory, information contained in the "most likely", "best case", or "worst case" fields should be free from WEPs because they are inherently probabilistic by definition. Yet, our participants relied on WEPs in many of their entries (Fig. 8).

[5] See Web App Developed by Dr. Joseph Ripberger from the University of Oklahoma for extensive bibliography of published literature on the subject.

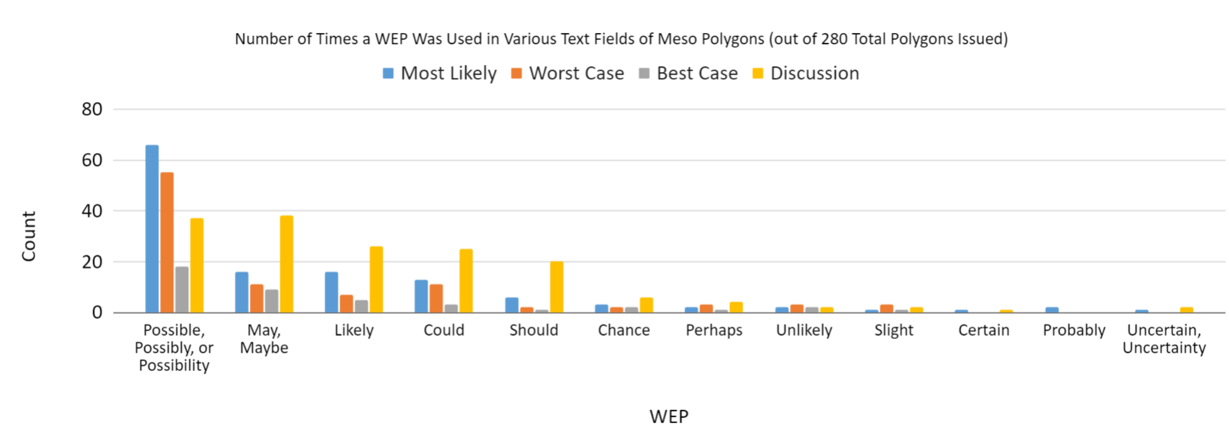

Figure 8: Histogram of the use of WEPs during the experiment, color coded by scenario or discussion.

To be clear, the OPG is not condemning the participants nor their entries. Instead, we are suggesting that the use of WEPs by our participants are representative of the larger body of forecasters. In fact, the OPG has gathered evidence showing forecasters frequently use WEPs in the AFDs. Using a script developed by Eric Allen of Eastern Region Headquarters, the OPG calculated the frequency of WEP use and percent of AFDs using WEPs issued from 2010 to 2020 (Fig. 9).

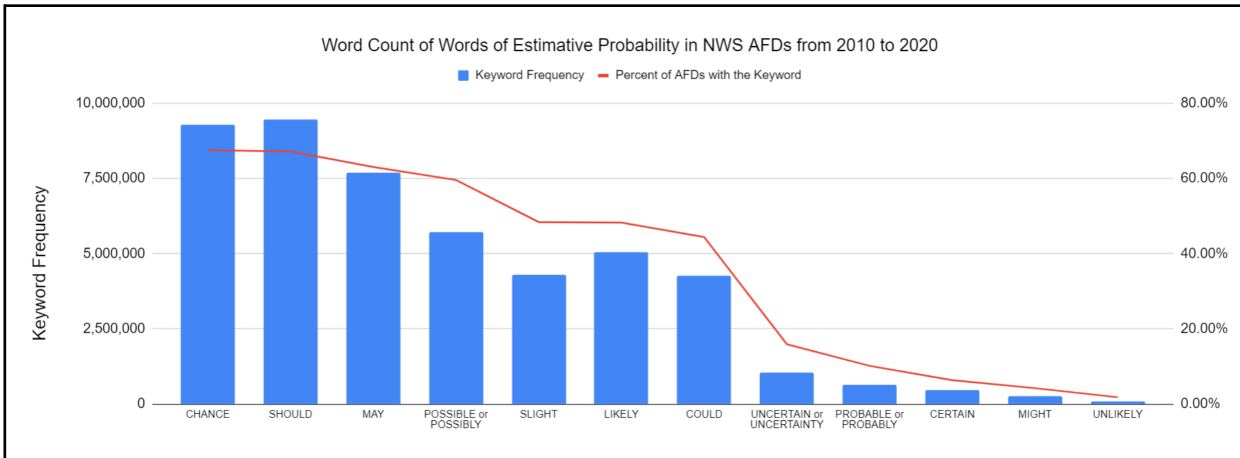

Figure 9: Histogram of WEPs within NWS AFDs issued between 2010 and 2020.

We do rely heavily on WEPs to describe uncertainty in one of our most common text products. This cultural norm[6] is simply manifested in participant responses in the mesoscale polygons created during our experiment.

[Dr. Ripberger's survey] showed that members of the public interpret these words in a highly variable manner. For example, the word "possible" or "possibly" can be interpreted as meaning anything from a

[6] Our participants served as a proxy control group for future experiments because they received no training (as part of the experiment) on communicating uncertainty nor did they review the results from Dr. Ripberger's work. The OPG will apply recommendations from Dr. Ripberger during FY22 experiments.

10% chance of occurrence to a 90% chance (similar results are noted for the words “may” and “chance” in Figures 10 and 11 below):

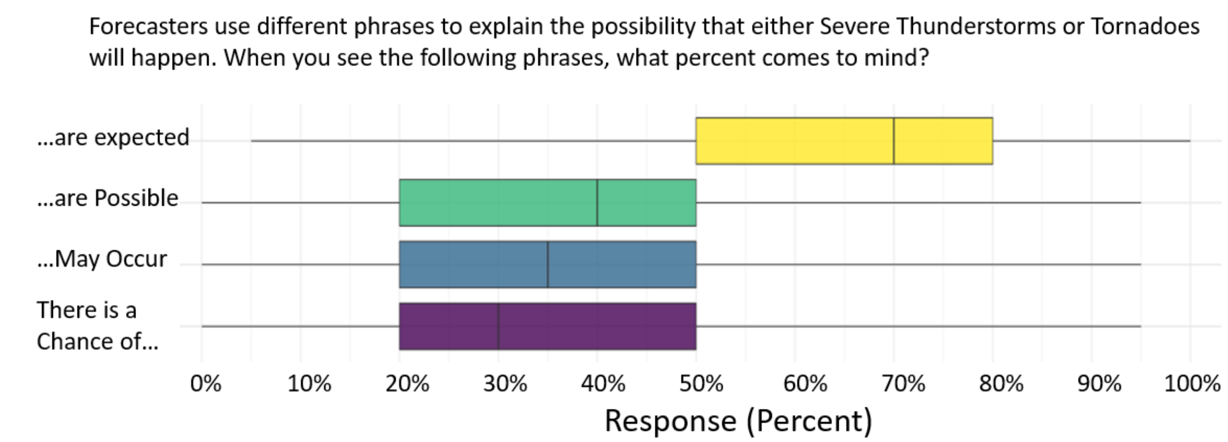

Figure 10: Box plot illustrating the range of probabilities associated with ‘no qualifying’ expressions to explain the possibility that severe thunderstorms or tornadoes may happen. Adapted from [Ripberger et al. (2021)](#).

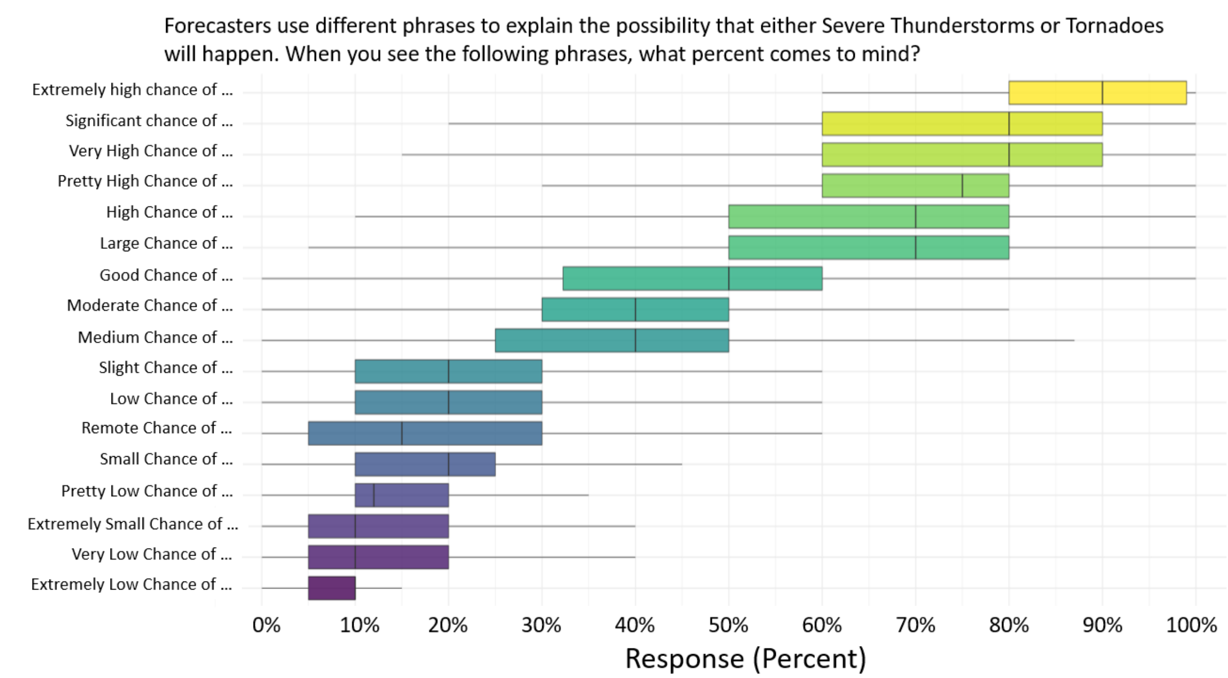

Figure 11: Box plot illustrating the range of probabilities associated with ‘qualifying’ expressions in terms of ‘chance’. Adapted from [Ripberger et al. (2021)](#).

The specific point of presenting this information is to note the challenge faced by the OPG in categorizing participant polygons when they used WEPs. The broader point is to suggest there is a more profound issue in the nature of risk communication leveraged by NWS forecasters in operations today.

4.2 Objective 2: CAMs and no-CAMs

Objective #2: Identify the methods that most effectively utilize both observational and model data to best anticipate severe hazards in the one to six hour time frame.

Spirit and Intent: The OPG understands that utilizing CAMs is ubiquitous in most forecast offices. However, we also understand that forecasters have expressed concern about "data overload." Moreover, participants in the OPG's Mesoanalyst Boot Camp in 2019 expressed how valuable observational data could be in aiding their ability to assess the mesoscale environment, even before analyzing CAM data. Thus, we want to begin the process of assessing the types of information that help maximize forecaster decision making. We understand that the end result will likely include a combination of data, but we believe we can better assess the relative contributions of each data source through experimentation in a safe environment.

The OPG believes there is abundant evidence from prior experiments, anecdotal conversations with forecasters, and prior operational experience suggesting there is simply too much information available to forecasters in operations[7]. Forecasters complain of "data overload" and we know from published research that "data overload" contributes to poor decision making:

> *"In some ways it was nice not having the CAMs, they are nice to look at but in so many cases they all have so many different solutions. They may hone in on one general area, but exactly how everything evolves, there can be so many differences, **sometimes it's an overload of data**. Which one do I believe? That one looks plausible, no, I could see that one, and **finally you just don't know what to do**."* (Emphasis added)

Bottom line, we believe the forecast process today is too complex in large part because of data overload. We believe the OPG can help identify the data, tools (visualizations), and workflows that improves the forecast process. Data denial experiments are crucial to this effort.

Hypothesis: Observational data and model data will both be essential to assessing the mesoscale environment and predicting short-term ACs. However, there will be particular time frames and situations within a convective event in which either observational data or model data will be more or less useful. Identification of these scenarios will aid in informing forecasters when to emphasize use of one dataset or another.

Evaluation Method: The OPG applied a data denial like approach by preventing the use of CAMs by one half of the participants. A second group of participants, operating in parallel with the first group, will incorporate CAM data into their decision making process. Each participant group will have access to

---

[7] It should be noted the data overload problem is likely worse for newer forecasters. Veteran forecasters have a greater likelihood of developing processes or workflows to mitigate the data overload problem.

SMEs during the experiment. The OPG will then use both a subjective and objective method for validating participant results:

- Objective: The OPG will compare participant polygons from each group to determine if there is a statistically significant difference between the CAM and no-CAM groups.
- Subjective: The OPG will collect participant feedback through a survey and group debriefing. We will ask questions associated with participant confidence, preferences, and perception of AC polygon quality.

NOTE: The OPG wants to be explicitly clear: we had no intention of arguing for or against the use of CAMs in the mesoanalysis process. We recognize the incredible importance of CAMs in short-term forecasting. Rather, our goal, as stated above, was to understand how different types of data influenced the forecast process and decision making.

4.2.1 Objective 2: Objective Results of CAMs and No-CAMs

The OPG found a statistically significant difference in the polygon lead times produced by each group. Interestingly, the difference is counterintuitive to our preconceived notions. The group with CAM data produced shorter lead times[8] than the group without CAMs. We also found that the polygon sizes were nearly statistically different[9] at the 95th confidence interval with the CAM group drawing larger polygons overall.

Otherwise, we found no statistical difference between the mesoscale polygons produced by the CAM and no-CAM groups in terms of polygon end lead times, durations, or verification (Fig 12).

[8] As a reminder, the term "lead time" in our experiment refers to the time between when the polygon was created and when the polygon first became valid. So, a polygon created at 4 pm that became valid at 5 pm would have a 60 min lead time.

[9] Technically, the Student's t-test p-value must be 0.05 or lower to be statistically significant.

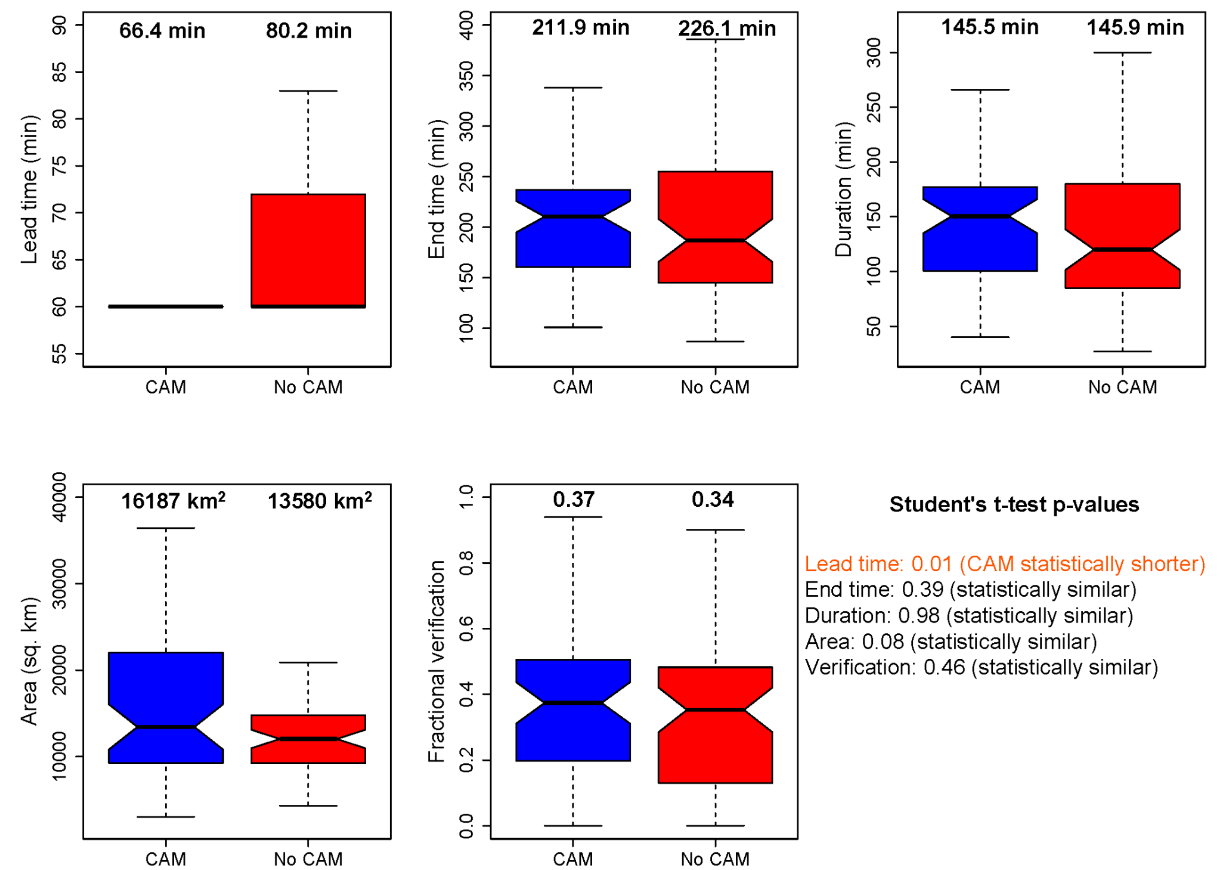

Figure 12: Notched box plots of: lead time, end time, duration, area, and fractional verification of the severe polygons issued by the CAM (blue) and no-CAM (red) groups. Above each boxplot is the mean of the distribution. Notches whose areas overlap indicate statistically similar medians. Student's t-test p-values are also provided in the bottom-right corner. Note that there are no notches for the lead time plot, as there was not enough spread in the CAM group lead time to accurately calculate them.

The OPG was quite surprised by these results. We expected to see the CAM group produce longer lead times, smaller polygons and superior verification.

Realistically though, given the relatively small sample size, and other variables in the experiment, we can not draw any firm conclusions from the results. We are certainly not suggesting CAMs offer no value. Rather, we believe our experimental design likely contributed to the unexpected results. In retrospect, we should have established more experimental constraints, provided greater pre-experiment training on various CAM data to establish a knowledge baseline, and run the experiment over many more events to increase our sample size and provide experience for our participants.

Regardless, we do believe we uncovered one potential root cause of our mixed verification results: data overload. Participants had so many different analysis options that eliminating one option, CAMs, had no meaningful effect on the verification. We are concerned that the current forecast process, and the resulting forecast accuracy, will be inherently limited until we objectively identify the data that promote the best decisions and eliminate the data that show no value.

We suggest our results highlight a further need to assess the data overload problem in operations. We propose an experiment, or set of experiments, that task forecasters with making predictions for various hazard scenarios and time frames. We would establish a baseline by limiting participants' access to very few data sets and verify forecasts for specific variables. Then, over time, we would introduce new data, verify predictions, and repeat until we have determined the right mix of data that maximize forecast skill.

Until then, we hope the reader finds the verification interesting, but takes no meaningful action from the results.

4.2.2 Objective 2: Subjective Results of CAMs and No-CAMs

During the experiment, the groups who were restricted from using CAMs often complained that they were missing a critical tool in their forecaster toolbox. In post experiment surveys and debriefings, these forecasters further articulated that CAMs directly support forecast confidence. CAMs (or CAM ensembles) help to either reinforce or refute conceptual models derived from observational analysis:

> *"The lack of CAMs was a hindrance. CAMs show you short-term evolutions that you can compare with the observations. It gives you a sense of confidence that the atmosphere is evolving based on what you assessed in the model data. You get info on storm mode from CAMs."*

Further, the "no-CAM" group noted that they struggled with crafting "most likely", "best case", and "worst case" information because they lacked the kind of objective and reliable probabilistic information that only comes from ensemble systems:

> *"By just having obs only heightened awareness of convective areas of interest and boundaries. But conceptualizing storm mode and storm intensity and potential scenarios made it difficult (without CAMs)."*

However, this desire for CAMs was not ubiquitous. Some participants stated that eliminating CAMs actually helped their process:

> *"I loved not using CAMs. I feel like we might rely on them too much. I've seen many times where they are behind and they are catching up. You are tapping into meteorological knowledge, this is what's going on and this is what I expect to happen without even looking at the CAMs."*

> *"It might be counterintuitive, but I almost felt more confident because I felt I could rely on the observations and the boundaries. My conceptual model that I developed to make the connection to… And then you see it happen, and then you get more confident in what you're thinking will happen next. It made me really want to use the CAMs less in mesoanalysis."*

Further, when asked, "*How effective do you feel you were at performing mesoanalysis (with CAMs or without CAMs)?*", both groups provided similar responses regarding their perception of effectiveness (Fig. 13).

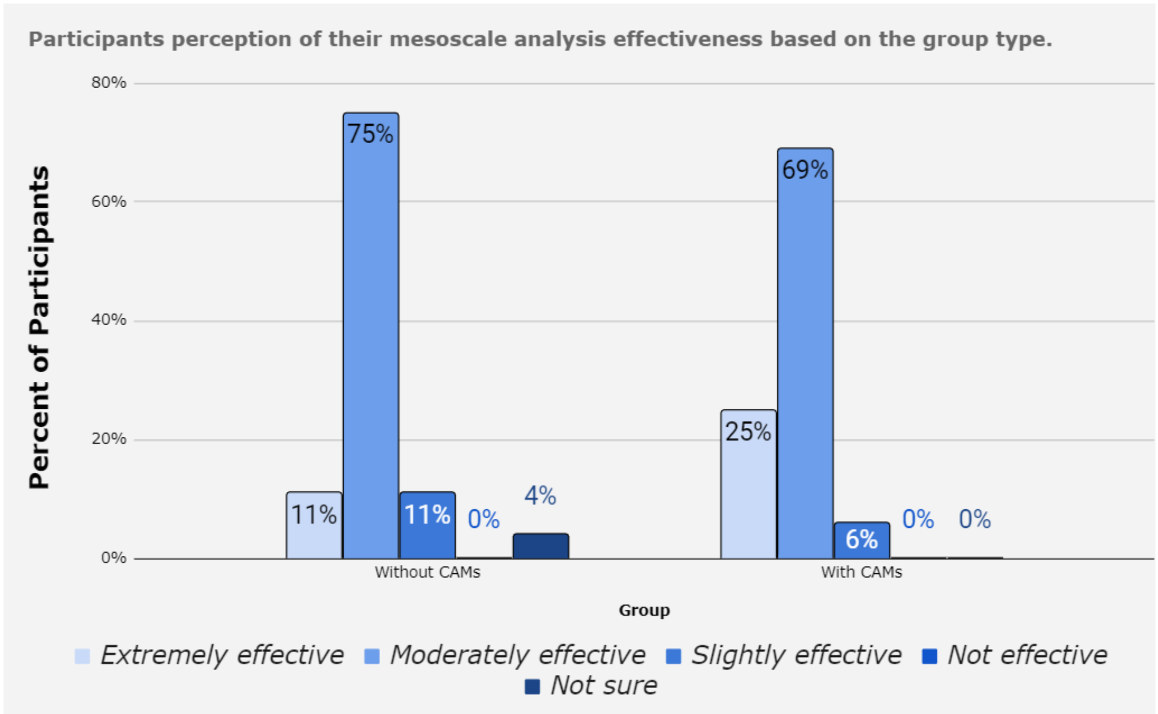

Figure 13: Participant perception of their effectiveness in conducting mesoanalysis within the no-CAM (left) and CAM (right) group, as measured by participant surveys.

While individuals in both groups felt they were effective at assessing the mesoscale environment and conveying that information, there was a clear distinction when we asked about confidence. The group who had access to CAMs felt far more confident in their ability to analyze the mesoscale environment compared to typical operations than the group without CAMs (Fig. 14):

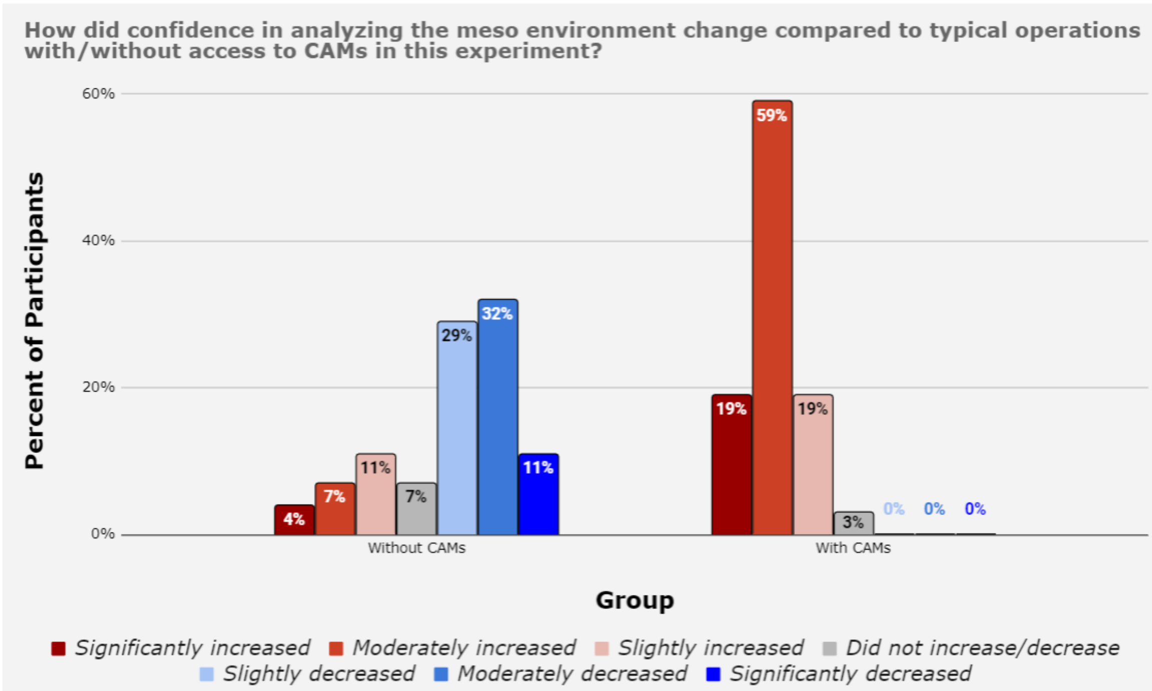

Figure 14: Participant perception of the impact of CAMs on their confidence in analyzing the mesoscale environment compared to operations, as measured by participant surveys. Increased confidence is shaded in red and decreased confidence is shaded in blue.

The group with access to CAMs stated that CAMs also helped provide valuable information regarding alternative scenarios:

> *"CAMs help you extrapolate the current environment forward. It helps build confidence."*

> *"Having CAMs gave insight into the options as the best and worst case. Even if you didn't believe one, it gave me an idea of what could be."*

Further, the group with CAM guidance felt far more comfortable responding to IDSS injects requesting information on exceeding certain thresholds. The OPG produced several injects that asked participants questions like, "*what are the chances we receive wind gusts above 40 mph?*", or "*how likely is it that we receive severe hail in the next few hours?*".

> *"The only time that I really kind of wish I had access to CAMs was when we had the inject about the probability of winds exceeding a threshold because I didn't want to just pull a random number out of my ear."*

One interesting caveat is that some forecasters would "throw out" or "lose track" of the CAM solutions if the model failed to initialize ongoing convection well.

> *"On our first day the CAMs weren't handling things well so we kept tossing out the CAMs and just focusing on the observations."*

"*I was focused on the obs and I lost track of the models sometimes.*"

The OPG would be interested to learn from model developers about the relationship between model initialization and downstream forecast accuracy. For example, many participants felt that if the HRRR failed to properly identify ongoing convection, then the entire model output should be considered "flawed" and rejected as a source of guidance. During the 2019 Mesoscale Bootcamp though, the OPG presented scenarios in which participants emphasis on initialization quality caused them to misdiagnose future events either because the model initialized well but still produced poor forecast guidance, or because the model initialized poorly but provided high quality guidance on timing and storm mode in future forecast hours.

Asking forecasters to refrain from using CAM guidance during the experiment produced interesting, but mixed, results. Drawing conclusions from these mixed results is therefore rather difficult (we could neither confirm nor refute our hypothesis). Thus, our finding is based on what we don't know rather than what we learned from our experiment.

4.3 Objective 3: Geospatial Collaboration Tools

The OPG would like to reiterate that our use of AGOL is intended to demonstrate collaboration capabilities for future operational tools. We are not advocating that AGOL become "Operational". AGOL is simply the most convenient, and admittedly powerful tool we have to meet our experiment objectives.

Objective #3: Evaluate the impact[10] of geospatial collaboration tools on the decision making process when assessing localized areas of potential severe hazards in the one to six hour forecast period.

Spirit and Intent: There are multiple ways in which collaboration can take place among meteorologists at neighboring WFOs (and potentially with NCEP centers). Often collaboration takes place via some form of chat in AWIPS Collaborate, in NWSchat, or in Google Chat. Conference calls are also a typical collaboration method in severe convective operations, especially with respect to watch collaboration. However, the OPG would like to evaluate the use of geospatial collaboration tools as we believe it may improve collaboration efficiency. Findings from prior OPG experiments involving collaboration have reinforced the importance of geospatial collaboration.

Hypothesis: The use of geospatial collaboration tools, such as AGOL, in conjunction with other standard tools, such as chat and video conferencing, will improve the collaboration experience. The focus on geospatial collaboration will improve forecasters skill in identifying ACs and increase forecaster confidence.

Evaluation Method: The OPG monitored portions of the collaboration process including: chat rooms, calls, and geospatial identification using AGOL and documented the relationship between collaboration and validation of the identified ACs, frequency of identifying those areas, and probability of severe

[10] The OPG understands that "impact" is subjective and may require additional experiments to evaluate the magnitude of the impact.

weather within the polygons. The OPG will also conduct a survey of the participants and a debriefing to more directly ascertain the effectiveness of aspects of the collaborative process.

Results: During prior OPG experiments focusing on collaboration, we identified strengths and weaknesses of various collaboration methods. For instance, we know chat based collaboration is quick and easy, but can lead to a "stream of consciousness" outcome. Participants felt that simple text chat, currently used in AWIPS collaboration, is too restrictive to effectively collaborate with others.

We know video conferencing greatly enhances the collaboration experience, but can be time consuming and exhausting. In addition, it can be challenging to determine who should participate in a video chat in real-world situations. Regardless, in both chat and video based collaboration cases, participants inevitably discuss "where" events may occur. The OPG has found that this geospatial component of collaboration is critical to success, but offices lack effective geospatial collaboration tools[11].

As a result, the OPG has investigated the use of AGOL[12] to demonstrate possible solutions to the geospatial collaboration challenge. Participants in our mesoanalysis experiment overwhelmingly felt the capabilities provided by AGOL were useful for the collaboration process (Fig. 15).

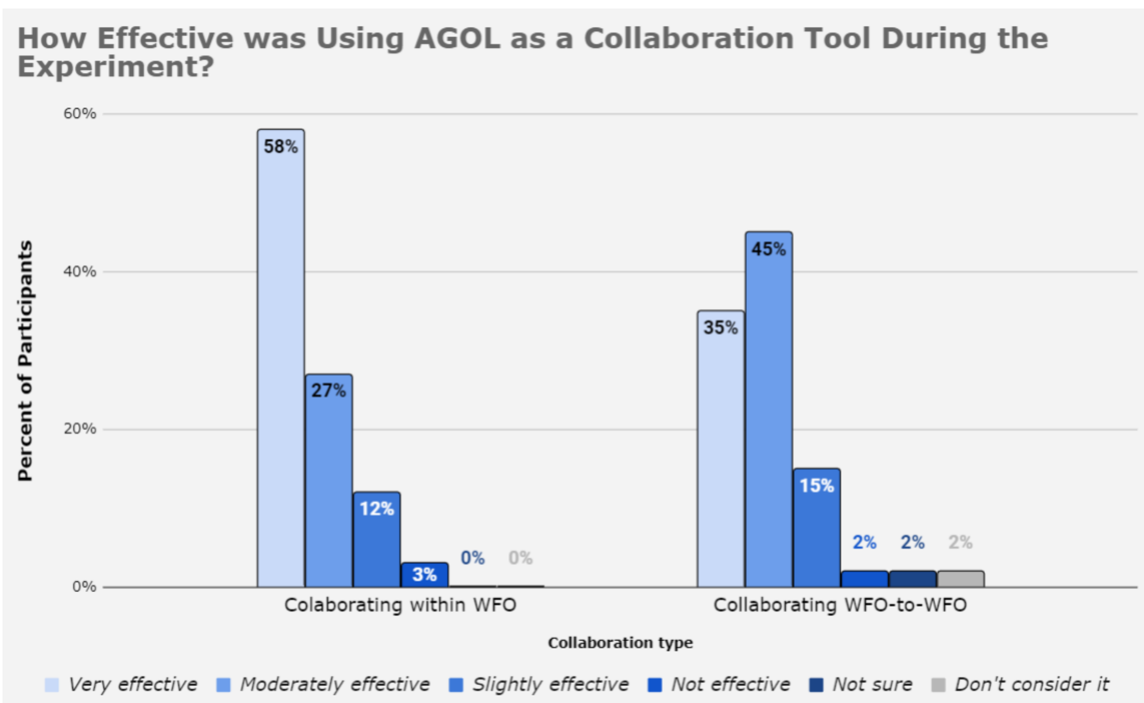

Figure 15: Participant perception of the effectiveness of AGOL as a geospatial collaboration tool, as measured by participant surveys.

"It was cool and helpful to see what other offices were keying in on (geospatially)."

[11] The general consensus from participants regarding mesoscale environment collaboration tools was that we need to use geospatial collaboration tools to maximize the effectiveness of collaboration.

[12] AGOL is available to all individuals within the NWS (and NOAA) and only requires an LDAP login for use.

"*I thought it worked really well. It was nice to see what other people were putting out*."

"*You could see your neighbor's polygons easily and read what they wrote and see what their thinking was and adjust accordingly even if I didn't formally reach out (via phone or chat).*"

"*I thought it was a really good platform. It would have made it more helpful if you could see the polygons before the final product was saved.*"

"*Most of our collaboration right now is in the grids (in GFE). GFE is not designed for the Mesoanalyst role.*"

"*I think that platform has serious potential. It was pretty exciting to use*."

Certainly there were suggestions for improvement, but the participants recognized the value of geospatial collaboration and felt this was a concept that should be integrated into operations in the future.

There are, of course, several side benefits of leveraging software like AGOL to support geospatial collaboration:

- All of the polygons and associated data are archived. After events, forecasters (or SOOs/DOHs) can review the information for verification or training purposes.
- Because we are capturing reasoning for the polygons in the discussion session, with time we can build a database detailing the kind of information that produced the most accurate outcomes/decisions.
- Because any person with a noaa.gov email address can access the maps and content, we can increase situational awareness of forecasting challenges from across the agency.
- The polygons could include an option for forecasters to express their desire for a collaboration video conference call. This would help national centers (NCEP) and Regional Operation Centers (ROCs) determine who should participate in a collaboration call and what topics are highest priority.

Finally, participants did bring up the difficulty of collaboration on IDSS. They noted that AGOL may not effectively support IDSS collaboration:

"The biggest thing was matching the polygons up from AGOL to the DSS graphic was kind of a tedious thing at first. Wanted to make sure they weren't highlighting 2 different areas."

However, participants also noted that IDSS collaboration is rather difficult today regardless of the tools. They pointed out how IDSS in operations is highly variable and inconsistent. Collaborating on content thus becomes a major challenge

4.4 Objectives 4 and 5: Communication Best Practices, and IDSS Communicator Role

Because Objectives 4 and 5 are so closely related, the two objectives are combined into a single section in this report. We have provided the objectives and their background first, followed by results and findings from both objectives.

Objective #4: Determine best practices of two communication chains during a rapidly evolving and dynamic situation:

1. **Internal collaboration between a WFO shift team**
2. **Collaboration between two (or more) WFOs**

Objective #5: Identify the feasible responsibilities of the IDSS Communicator working in tandem with the dedicated Mesoanalyst.

Spirit and Intent: In an operational work environment, collaboration occurs both within an office (e.g. among shift team members) and with neighboring offices. Building off objective #3, we attempted to identify best practices to improve the efficiency of collaboration during a real time, dynamic, and rapidly evolving event.

Further, we wanted to assess the viability and value of assigning the role of "IDSS Communicator" to one of the participants in each simulated shift team. We intended to demonstrate a streamlined process that provides a continuous flow of information during real-time evolving events; the basis of the FACETs initiative. By doing so, the OPG established a third chain of communication: between the local WFO and Core Partners[13] (IDSS), further supporting objective #4. The OPG also examined the workload of the IDSS Communicator in relation to the workload of the Mesoanalyst.

Hypothesis 1: By establishing clear roles and responsibilities, providing a structured approach to collaboration, and communication best practices, we will achieve an effective and rewarding collaboration experience for our participants.

Hypothesis 2: While the IDSS Communicator may at times collaborate with the dedicated Mesoanalyst as an "assistant Mesoanalyst," their primary role is to use the information provided by the Mesoanalyst to create IDSS content. The IDSS Communicator can reasonably provide frequent IDSS content in the form of IDSS graphics, posts to social media, and answering partner questions.

Evaluation Methods: The OPG monitored the collaboration process (chat rooms, video calls, IDSS content, and AGOL output) then conducted a survey and debriefing to more directly ascertain the effectiveness of aspects of the collaborative process.

[13] We did not include real core partners in our experiment this time. Instead, OPG members assumed the role of Core Partner when needed.

Results: To understand the impact of the communication structure provided by the OPG for this experiment, we asked participants to describe typical collaborative experience in real operations. Then, we asked whether or not participants felt the OPG experience was more or less effective than in operations.

Over 50% of the participants said they "seldom" to "never" collaborate on the mesoscale environment in current operations (Fig. 16).

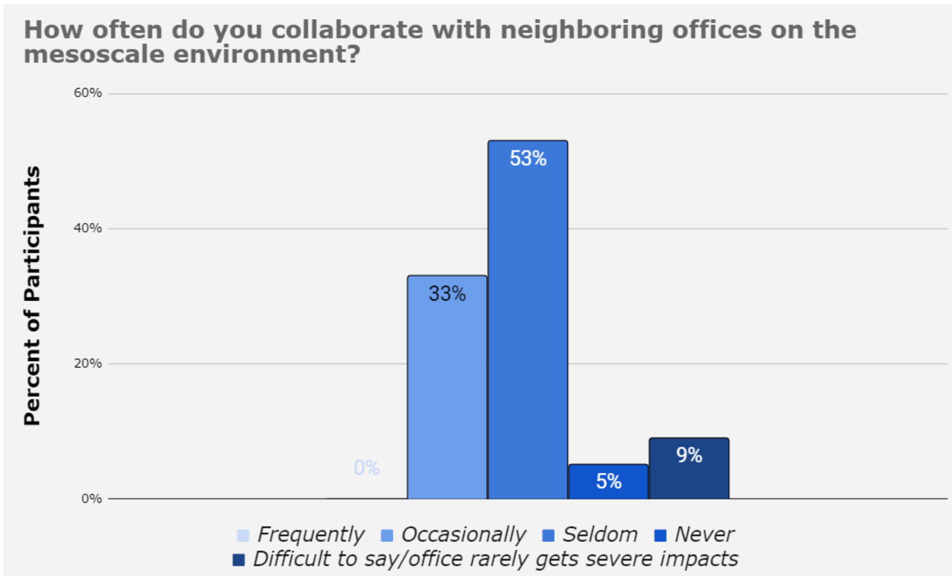

Figure 16: Histogram of the frequency of collaboration with neighboring offices regarding the mesoscale environment in current operations.

One could argue that the lack of mesoscale collaboration makes sense given the potential for mesoscale events to exist entirely within a CWA. However, overwhelming evidence from this and prior experiments shows that collaboration with fellow meteorologists improves forecaster confidence, forecast consistency, and (in theory[14]) forecast accuracy. Further, several regional tests focusing on mesoscale collaboration have occurred in both Central and Southern Regions providing even greater support for the concept.

When participants did conduct collaboration in operations for mesoscale events, they noted the experience was slightly to moderately effective nearly 60% of the time. Interestingly, twice as many participants felt their mesoscale collaborative experiences in current operations were not effective than were very effective (Fig. 17).

[14] The OPG has not formally tested the difference in forecast accuracy between collaborative and non-collaborative forecasts. However, evidence from the book "Superforecasting'' by Tetlock and Gardner suggests collaboration leads to greater accuracy - a concept the OPG intends to formally test in the future.

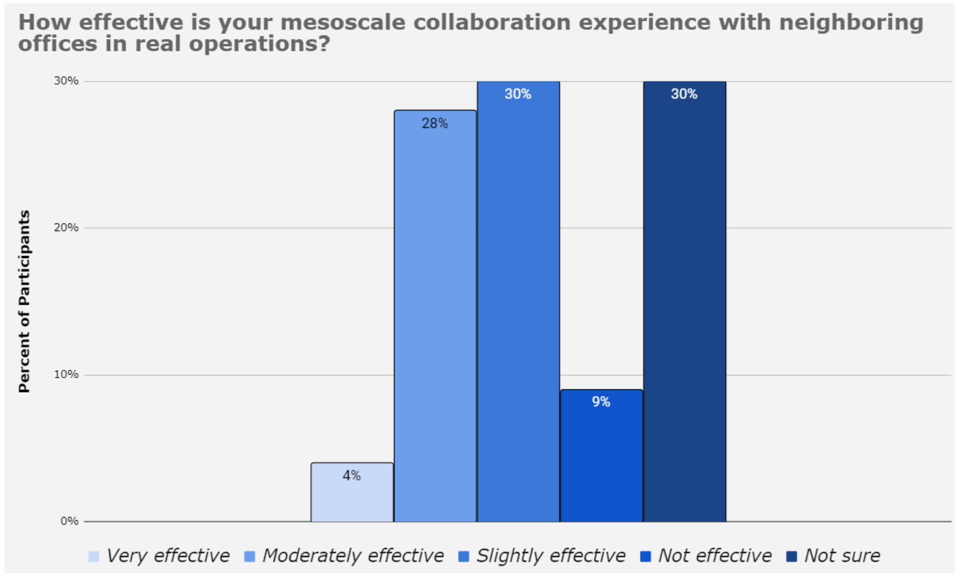

Figure 17: Participant perception of the effectiveness of collaboration with neighboring offices, when it does happen, in current operations.

The key takeaway from this is that mesoscale collaboration in operations seldom happens and when it does, there is no guarantee that the experience will be on the good side of the effectiveness scale.

During our experiment though, our participants engaged in constant collaboration via Google Meets with their fellow WFO shift team members and their assigned SME. Over 90% of participants in both the CAM or no-CAM group felt they were moderately to very effective at collaborating on the mesoscale environment within their simulated WFO.

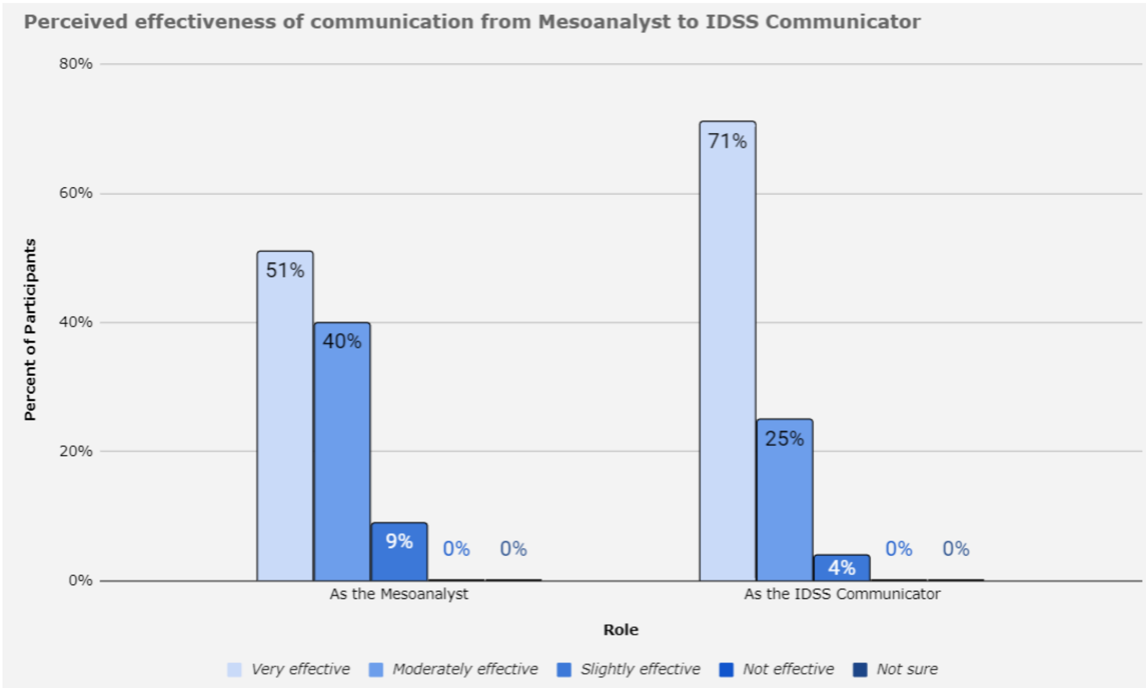

Figure 18: Participant perception of the effectiveness of communication from the Mesoanalyst to the IDSS Communicator, separated by roles (as Mesoanalyst on the left and as the IDSS communicator on the right).

Further, we noted in figure 13 in section 4.2.2 above that over 75% of our participants felt they were moderately to very effective at analyzing the mesoscale environment: a fact that is directly related to their collaborative experience.

Coincidentally, our participants collaborated with neighboring WFOs during the experiment at about the same rate they do in operations today (Fig. 19).

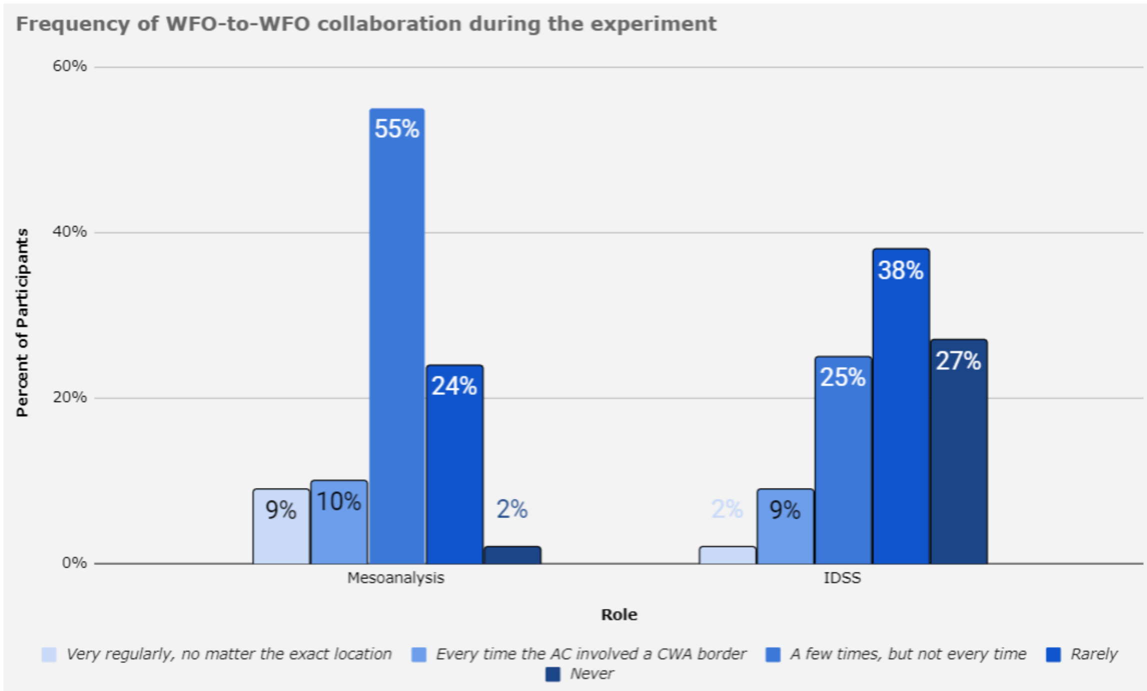

Figure 19: Frequency of WFO-to-WFO (neighboring office) collaboration during the experiment on mesoanalysis (left) or IDSS content (right).

Finally, forecasters noted their confidence generally increased when inter office collaboration did occur during our experiment (Fig. 20).

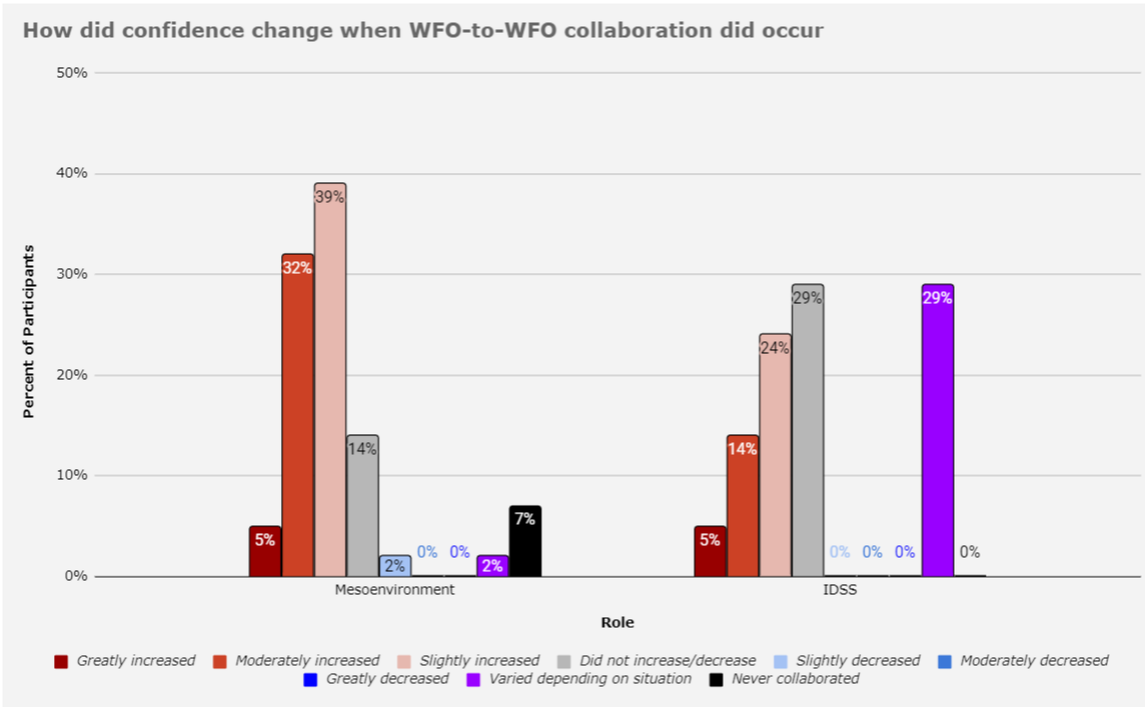

Figure 20: Participant perception of the impact of collaboration on confidence in analyzing the mesoscale environment (left) or producing IDSS (right). Increased confidence is shaded in red and decreased confidence is shaded in blue.

The survey results showed that during our experiment, forecasters felt they were effective at assessing the mesoscale environment, they collaborated effectively, and their confidence increased thanks in large part to various collaboration efforts.

Additionally, asking participants to adopt specific roles generally helped streamline the workflow and collaborative process:

> *"The Mesoanalyst fed info to the IDSS communicator and that was effective. Let the Mesoanalyst focus on the environment and then inform the IDSS person on where to target messaging."*

> *"In the groups that I worked in we tended to do things very collaboratively. Even though in the end one person was producing (the polygon) and another person was producing (IDSS content). The whole mesoanalysis itself was very collaborative."*

> *"Assigned duties help keep anything from really sliding through the cracks, but also helps put people where they are the best so that they can thrive. Having assigned duties and roles helps communication flow. It builds communication."*

Although, to be fair, some participants felt the workload fell more on the dedicated Mesoanalyst than the IDSS communicator:

"I was focused on the obs and I lost track of the models sometimes. I felt like the Mesoanalyst had more to do than the IDSS person. I felt like I was kind of waiting as the IDSS person."

Others felt the roles and responsibilities break down during the course of an event:

"It happens in a real WFO that the roles are organic. Starting roles are good, but those roles often evolve during a shift."

In designing this experiment, the OPG intended participants to work collaboratively on the mesoanalysis, but split their final tasks between the shift team. We may have failed to communicate that effectively, but most participants fell into that workflow and realized the value:

"It can be helpful to still be really tuned into the mesoanalysis so that you can create the messaging without constantly asking the Meosanalyst to summarize what they are thinking."

"There was a lot of overlap in the roles. Both of us were doing mesoanalysis. You had to have a shared workload and bounce ideas off each other."

"Even as the IDSS person, you are doing meso....The only separation was the individual tasks."

"I think it's critical that you have everything defined beforehand. You can manipulate as you move on, but you gotta have it defined before you start otherwise it's a mess. I don't see how you can possibly do a decent job without having your roles defined."

4.5 Objectives 6 and 7: Evaluating Probabilistic Forecasting[15]

Probabilistic thinking is the future, but the road to get there is complicated and requires analysis of both transmission (NWS) and reception (core partners) of probabilistic data.

We know local office forecasters are often asked by core partners to provide a "worst case scenario". Yet, anecdotal evidence suggests forecasters typically use a deterministic process when designing the forecast[16].

"Not every forecaster is attuned to probabilistic thinking when it comes to convective threats. Many remain in the binary mode of severe/no severe."

By producing "most likely", "best case", and "worst case" scenarios as part of the experiment, we were effectively asking participants to generate a Probability Density Function (PDF) curve of the magnitude of sensible weather conditions.

---

[15] Similar to Objectives 4 and 5, Objectives 6 and 7 are very closely related so we have combined them into a single section.

[16] This trend is definitely changing. Matt Jeglum at WRH recently produced evidence showing that the use of the words "ensemble", "GEFS", and "EPS - ECMWF Ensemble" have greatly increased in AFDs in the last couple years.

The OPG had no intention of assessing participants' skill or reliability in predicting best case or worst case conditions. Rather, we designed the experiment in part to assess the impact of a probabilistic mindset on the short term forecast process as stated in Objectives 6 and 7.

Objective #6: Determine the ability for the Mesoanalyst to explicitly provide a best case (10th percentile) and worst case (90th percentile) scenario in real time in various situations, given the datasets available to them.

Objective #7: Assess the impact of explicitly determining the best case and worst case scenarios on: 1) the Mesoanalyst's ability to shift their most-likely expectation when necessary, and 2) the IDSS provided throughout the event.

Spirit and Intent: The OPG wanted to continue efforts from the 2019 Mesoscale Bootcamp where we asked participants to describe the PDF curve of various events. We attempted to visualize, to some degree, the probability of severe weather developing within the ACs in real time. We hoped to build an understanding of how probabilistic thinking impacted their forecast process.

The OPG also wanted to understand the ability for the Mesoanalyst to reassess and reevaluate their scenarios (best, likely, worst) under different circumstances and dynamically evolving conditions. We also wanted to understand how the IDSS Communicator uses probabilistic information provided by the Mesoanalyst.

Hypothesis 1: A Mesoanalyst may, at times, find it difficult to communicate the reasonable best case outcome or the reasonable worst case outcome, or both, depending on the specific mesoscale situation and the data sources and data output available to them. However, conducting such an effort should improve the Mesoanalysts understanding of the range of possible outcomes.

Hypothesis 2: A previous determination of the range of possible severe hazards by the Mesoanalyst will allow the Mesoanalyst to more easily shift what they determine to be the most likely scenario when/if it is necessary to do so later in the event. The IDSS Communicator will be able to more readily convey a range of possible severe hazards, including messaging the reasonable best or worst case scenarios.

Evaluation Method: The OPG surveyed the participants to determine how comfortable they were at explicitly determining reasonable best case and reasonable worst case outcomes for each polygon. We also investigated content participants provided in the "best case" and "worst case" polygon data fields to identify trends. Further, we discussed their perspectives on shifting their most-likely outcome when necessary as the mesoscale environment evolved.

Results: Without question, asking participants to provide best case, worst case, and most likely information during the experiment proved to be the most challenging:

> *"It is often challenging to determine an appropriate best and worst case scenario, especially in low confidence events (which is often the norm in my area). It is also particularly challenging to do it consistently from forecaster to forecaster."*

The groups with access to CAMs were more comfortable with the exercise as they felt the CAMs provided explicit probability information.

> *"(CAM ensembles)...can increase or decrease your expectations for best, most likely, and worst case scenarios...**even seemingly bad model runs can provide useful information in terms of depicting the potential worst case scenarios, spectrum of timing, and storm mode**."* Emphasis added

Further, participants expressed conflicting views on the value of creating and sharing scenario based information.

> *"Perhaps it's a best practice to build your meso conceptual model but I am not sure explicitly stating it or documenting it has an operational benefit at least internally."*

Some also expressed concern that partners may gravitate to the worst case scenario automatically, which can be problematic. Further, external entities may unfairly criticize NWS forecasts by "cherry-picking" only portions of the scenario based messaging they felt were least accurate. Or, members of the public may infer scenarios to imply that, *"we (the NWS) don't know what's going to happen."*

The common theme from those who had reservations about scenario based messaging involved supposition rather than object observation though. They spoke in terms of how they felt, or how they believed partners may respond, but there were no concrete examples of how scenario based messaging negatively affected some real event.

On the contrary, one participant did recount a story from the experiment that reinforces the idea of thinking probabilistically and scenario based:

> *"Being honest: when first introduced to these categories, my first thought was 'This would be considered time consuming for a Mesoanalyst for [an] impactful convective day, and likely would be ignored in operations.' It just so happens, I was a part of the team that covered Goodland for May 24th (during the experiment), which included multiple tornadoes/flash flooding across the CWA. In this case, the event definitely lived up to the "worst case" scenario that was being analyzed during the mesoanalysis. **Without doing the best/worst case analysis, I don't know if that would have been communicated as being an anticipated outcome. Needless to say, that made a believer out of me as being important.**"* (Emphasis added)

Overall, 93% (56 out of 60 participants) said that explicitly stating the best, most-likely, and worst case scenarios among fellow forecasters at their home WFO (i.e., not with partners) regularly added some value during a potentially impactful convective event (Fig. 21).

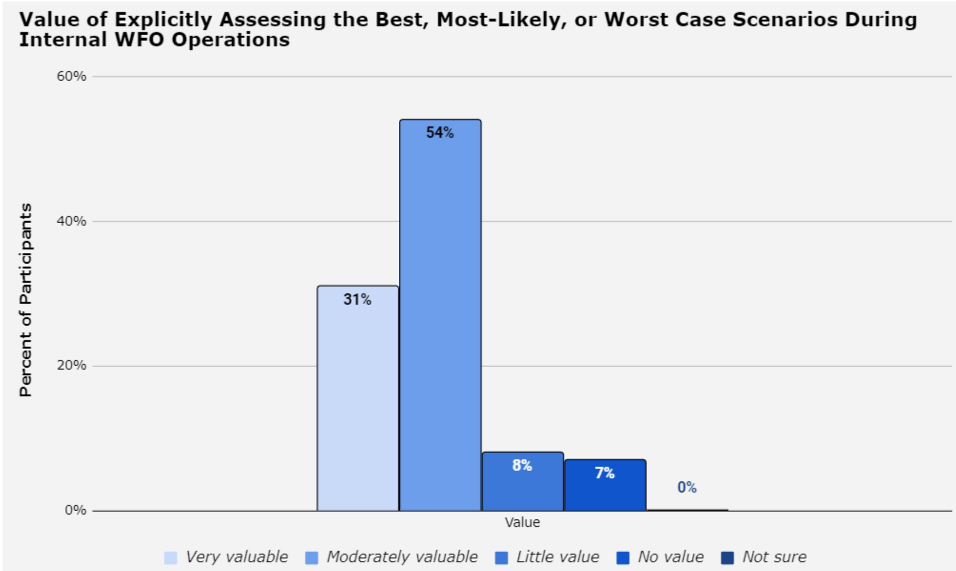

Figure 21: Participant perceptions of the 'value' that explicitly stating probabilistic scenarios has on internal WFO operations.

Further, participants felt comfortable both creating the AC polygons and assigning scenarios (although they were less comfortable with assigning scenarios; Fig. 22).

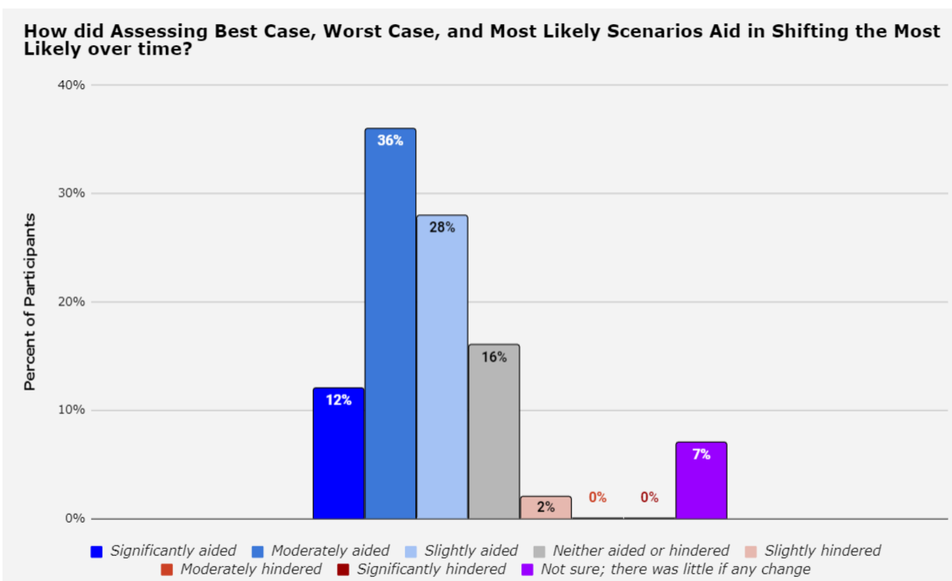

Figure 22: Participant comfortability of creating ACs (left) and assigning scenarios (right) during the experiment, as indicated by survey responses. Uncomfortable is shaded red and comfortable is shaded blue.

Participants overwhelmingly felt that assessing scenarios increased their confidence when communicating IDSS information (Fig. 23).

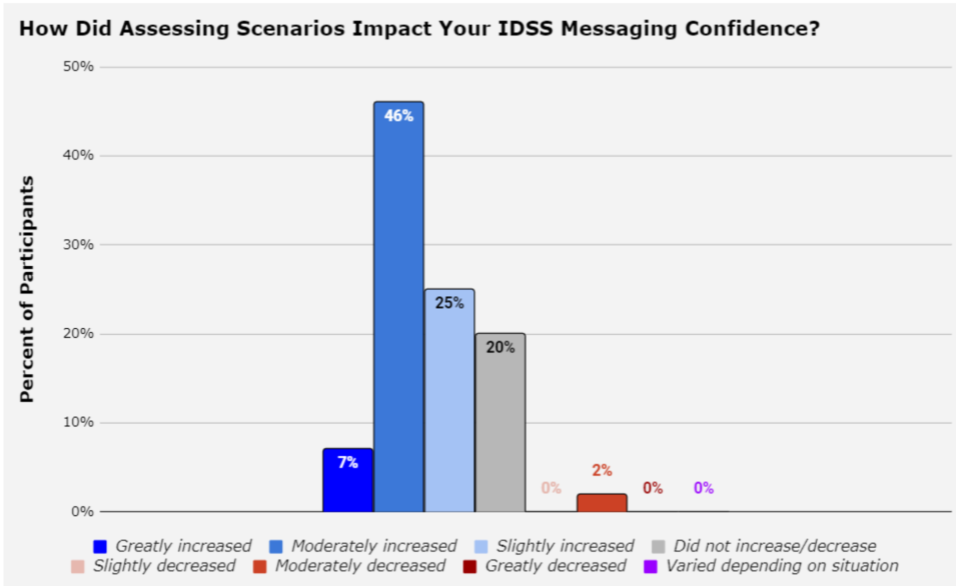

Figure 23: Participant perception of impact of assessing scenarios on confidence in IDSS messaging.Increased confidence is shaded in blue and decreased confidence is shaded red.

Finally, participants stated that assessing scenarios helped them shift the most likely conditions over time as an event evolved (Fig. 24).

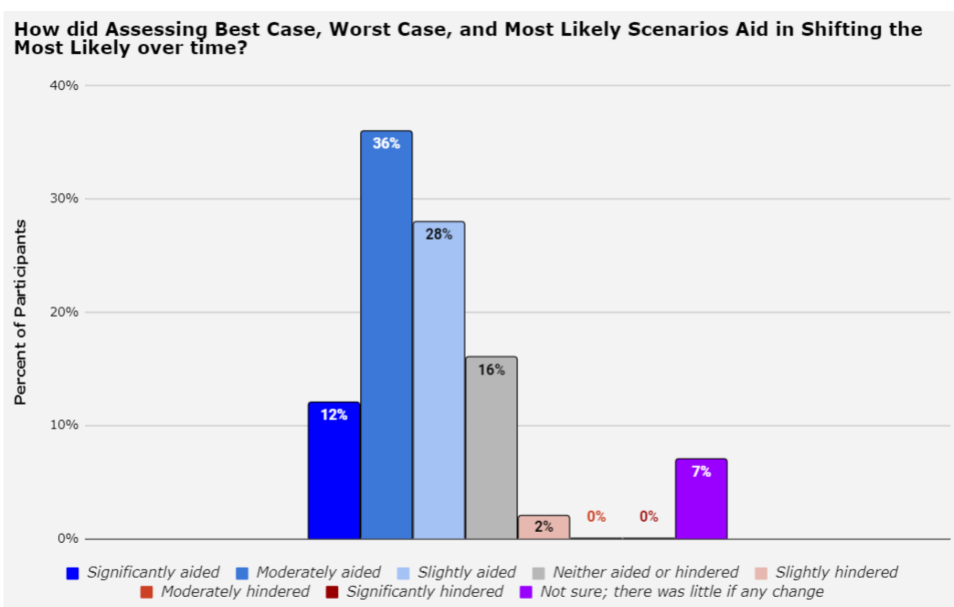

Figure 24: Participant perception on the impact of scenario assessment on the Most Likely case.

5. Non-Objective Related Findings

As with all OPG experiments, we occasionally uncover important information that was not necessarily within the experimental design. In this case, there were two findings that the OPG feels are important to share with the reader.

5.1 Lack of Mesoanalysis In Operations

First, the overwhelming sentiment of participants in this and prior mesoscale related experiments (and indeed the general sentiment of field offices) is that mesoanalysis is critically important to operations. Further, skillful mesoanalysis is necessary in providing real time IDSS during dynamically changing events.

Through anecdotal conversations with local offices and several quotes from experiment participants, the OPG believes that local offices are highly inconsistent in their use of a dedicated mesoscale analyst.

During the debriefings from the mesoanalysis experiment, several participants noted how the mesoanalysis role in real operations was either left unfilled during severe events, or would be repurposed for other activities:

> "*The Mesoanalyst seems to often get pulled into other tasks in real operations, most times that is related to staffing. As soon as someone gets pulled into that other task, the game is then off for mesoanalysis.*"

> "*I have been assigned other duties while serving as the Mesoanalyst and it didn't work out very well.*"

> "*It can be easy to be thrown a curve ball because it's not the busiest of positions so you look like you're not doing much. But then it's hard to get back to your role*."

The OPG wanted to better understand why the Mesoanalyst role often falls off the plate in operations. Participants noted three common reasons: lack of staffing to fill the role (short staffed), lack of training on providing effective mesoanalysis, and culture (some people do not see the value).

Fortunately, training is now widely available through the Mesoscale Environment Awareness courses (among others), and many offices are now conducting focused seminars, group discussions, and WES cases on mesoanalysis.

Further, the OPG believes that results from experiments like this (and others from various testbeds), case studies, and real operational experiences will continue to demonstrate the value of mesoanalysis. In theory, providing proof in various formats will help address any lingering cultural concerns.

That leaves the issue of staffing. Certainly the OPG understands that an office cannot fill a role if they lack the people. However, there is a plausible alternative explanation for the issue of limited resources: operational inefficiency. Indeed, participants during the experiment provided clear statements on the issue:

"*Anything to cut down on redundancy. Personally I feel like we do a lot of the same things in different formats, whether that is an AFD, SPS, Warnings, Watches, Social Media graphics. It's all rooted in the same information. Cutting down on that would be helpful from an operational perspective. Not sure if that's the right answer since everyone gets information in different ways. But it makes it more efficient on our end.*"

"*I have to agree [that we sacrifice something like mesoanalysis because of these inefficiencies].*"

"*Amen. During active weather we are writing an AFD! Then we make briefing slides, then we use the briefing to make social media graphics. We can streamline this. If there was a way that we had some unified tool system where we can draw something and it would send it out to the rest of the world, and eliminate having to go in and post on Facebook, and post on Twitter, and even post on Instagram in some offices, that would help our workflow immensely.*"

The OPG believes that these redundancies, and the associated workload, effectively explain why offices must often relegate the mesoanalysis role. To make matters worse, these redundancies are duplicated across multiple offices in various events. In short, many people (often in multiple locations) are translating the same information into many different formats for the same event.

Creating some of the content (like a tweet) requires very little time of any individual, and some content leverages automation (like PowerPoint macros), but the culmination of all of these efforts creates a level of inefficiency that the OPG believes is a root cause for lack of widespread adoption and application of mesoanalysis.

6. Summary of Findings and Recommendations

Finding 1: We found the mutual aid approach in support of local office mesoscale analysis worked very well. Objectively, we found that the SMEs improved the predictive skill of our participants. Subjectively, we found the SME provided valuable insights and structure that increased participant confidence regardless of the participants prior experience.

Recommendation 1: To maximize the effectiveness of local office mesoscale analysis and IDSS, we recommend offices leverage subject matter experts through a “Mutual Aid” approach.

Finding 2: We found that Cloud AWIPS, and other cloud technologies such as AGOL, unlock untapped potential in support of operations. We found that forecasters can effectively analyze the mesoscale environment from home (remotely) using Cloud AWIPS (and other web based tools) and provide useful collaborative geospatial information to fellow forecasters to use in IDSS or warning decision making. We found that geospatial collaboration tools greatly enhanced and streamlined the collaboration experience by reducing spatial ambiguity and producing a focused deliverable.

Recommendation 2: To maximize the effectiveness of the mesoscale collaborative experience, we recommend two methods (can be integrated into operations today, or tested on larger scales):

1. Utilize geospatial collaboration tools to streamline the collaboration process and reduce spatial ambiguity
2. Produce a focused deliverable[17] with actionable information to forecasters that highlight the time and location where severe convection, or new convection, or improving conditions are anticipated

Finding 3: We found that our forecasters, in both our experiment and real operations, rely on “Words of Estimative Probability” (WEPs) when discussing uncertainty. We know from several published papers across multiple disciplines that WEPs lead to confusion and often result in poor decision making. In our case, OPG staff members struggled to assess the nature of participant predictions. In the real world, this confusion can lead to poor decision making during high impact events.

Recommendation 3: The NWS should consider conducting a thorough evaluation of official NWS products and IDSS content to understand the extent of WEP use by forecasters. The NWS should ensure forecasters understand the dangers of using WEPs by providing training or access to research compiled by Dr. Ripberger. Finally, the NWS should design (with the aid of social scientists) and test clear communication methods or strategies free of WEPs when discussing uncertainty in any NWS product or service.

[17] By deliverable, we mean a tangible bit of information intended to help forecasters make informed warnings or IDSS decisions. We are NOT talking about a public or partner “product”.

Finding 4: The overwhelming position of our participants from several mesoscale related OPG activities is that mesoscale analysis is crucial to forecast operations regardless of the hazard. Yet we know from both participant feedback, experiential, and anecdotal evidence that we do not have a consistent use of mesoscale analysis in local operations today. Participants have provided several reasons for not filing a Mesoanalyst role in operations:

- Lack of training
- Lack of staffing
- Operational culture
- Lack of experience/expertise
- Lack of time

The OPG believes the NWS has several effective solutions for most of the reasons listed above. We now have a robust web based mesoscale analysis training course available, we can supplement the staffing concerns with a mutual aid approach, we can influence culture over time by showing value in experiments and real operations, and the OPG and OCLO can generate experience through real-time demonstrations using Cloud AWIPS. However, participants still face a "lack of time" problem.

Based on information provided by participants in this and prior experiments, the OPG has hypothesized two plausible root causes for the "lack of time" problem: operational inefficiencies and data overload.

When forecasters have to repackage the same information in multiple different formats, often manually, they simply run out of time to effectively analyze the mesoscale environment. Further, when these same forecasters are faced with a data overload problem, they require time to sift through loads of information to draw meaningful conclusions about the atmosphere. In short, the OPG believes that a streamlined approach to forecast operations could potentially reduce the "lack of time" problem.

Recommendation 4: The NWS should consider conducting a thorough evaluation of operational activities to identify inefficiencies, redundancies, and potential service gaps. Further, the NWS should consider supporting data denial experiments to identify the data types and visualizations that maximize forecaster decision making while reducing noise (data that shows no value in forecaster decision making). Finally, the NWS should consider developing and conducting experiments in Test Beds and/or the OPG that focus on operational efficiencies, reduced redundancy and collaborative decision making.

7. Thank You

On behalf of the OPG, I would like to thank our participants and our SMEs for their incredible support and effort before and during the experiment. You performed exceptionally well - without question. My entire team at the OPG did not know what to expect in bringing together individuals with diverse backgrounds, locations, and experiences. We were nervous, but you once again proved why NWS forecasters are the best of the best. You taught us lessons about teamwork and mesoscale analysis techniques that will undoubtedly help the entire agency.

Thank you for your time and your trust in the OPG. We hope you enjoyed your experience and are able to extend the lessons you learned with us to your colleagues. We also hope you are willing to join in future OPG activities.

And to my team at the OPG: you are incredible individuals who are helping show what the future holds. Thank you for everything.

JJ

"When something is important enough you do it, even if the odds are not in your favor."
E. Musk